

nature



TRIANGULAR MOMENTUM

Topological effects give electrically driven laser the edge

Border crossing

Premature US deaths linked to air pollution from next-door states

Bronchial boost

Lung cells of ex-smokers show reduction in mutations

Promiscuous protein

Secrets of a malaria parasite's sugar transporter

Vol. 578, No. 7784
nature.com

Make universal health care a priority

World leaders and international donors must help to strengthen the health systems of the most vulnerable nations.

As the 2019 novel coronavirus continues its deadly rampage, the World Health Organization (WHO) is rightly drawing attention to the risks the virus poses to the poorest and most vulnerable nations – particularly in Africa.

As *Nature* went to press, more than 43,000 infections and more than 1,000 deaths had been confirmed. Soon, thousands of China's citizens will be returning to their jobs on the African continent after an extended new-year holiday. If the virus also reaches Africa, it could spread rapidly and undetected because health systems in many regions are too fragile and underfunded to cope.

As a result, the WHO has scrambled to equip 14 countries – including the Democratic Republic of the Congo, Ethiopia and Nigeria – with diagnostics, expertise and equipment to detect and contain the virus. The agency has also appealed for US\$675 million to assist vulnerable countries – an amount that it estimates will last only until the end of April.

And yet, as donors start to provide emergency aid – the Bill & Melinda Gates Foundation was among the first with a \$100-million pledge – it's hard to avoid the feeling of déjà vu. Infectious-disease outbreaks are often accompanied by such pledges to improve disease surveillance, and by promises to provide funds for drug and vaccine development. What is less forthcoming is sustainable funding for clinics providing community-level general medicine, and for medical and nursing education, as well as investments to sustain hospitals with supplies, electricity and running water.

These are all steps that would help countries to combat infectious diseases and improve overall public health – as WHO director-general Tedros Adhanom Ghebreyesus urged in a statement at the end of last month. Seven of the nations that the WHO will be helping scarcely have one nurse per 1,000 people, according to the most recent statistics from the World Bank. And more than 50% of the continent's 1.2 billion inhabitants lack access to essential primary care.

To be fair, a shift in outlook has already begun. In 2016, the World Bank and the Global Fund to Fight AIDS, Tuberculosis and Malaria committed \$24 billion over three to five years for universal health care in Africa. And Rwanda's president, Paul Kagame, is leading an African Union task force to achieve measurable universal health coverage in all of its 55 member states, partly by committing to spending 5% of gross domestic product on health care.

“
Seven of the nations that the WHO will be helping scarcely have one nurse per 1,000 people.”

A temporary surge of assistance aimed at infectious-disease surveillance – as is happening now – might suffice in places where health systems are reasonably robust. But for the poorest countries with the weakest systems, even the best projects will struggle once these grants come to an end, as the case of Ebola shows all too well.

After the world's biggest Ebola outbreak ended in 2016, donors, including the US government and the World Bank, put more than \$100 million into initiatives to strengthen health and disease-surveillance systems in the three countries that were worst hit – Liberia, Sierra Leone and Guinea.

But many of these initiatives are ending, and health care is showing signs of erosion. Since last summer, protests have been erupting in Liberia as the economy and the national health system have crumbled. Major hospitals are reported to lack life-saving drugs, and health workers and lab technicians say they have not been paid for months. Patients have been turned away from clinics empty-handed. This problem isn't specific to Liberia. In many of the poorest countries, staff in national health systems barely earn a living.

International donors have reasons for not providing long-term funding for salaries for public employees. One of their biggest fears is that in doing so they would become too deeply involved in the workings of government departments, which are often complicated organizations to navigate. Another worry is that donors could be perceived as telling sovereign governments what to do.

Clearly, finding solutions to these problems will not be easy, but donors must consider how their initiatives can help to strengthen national health systems for the long term. For example, they could ensure that the health workers being trained to handle patients suspected of having coronavirus are still employed at hospitals five years later. This might not seem like a priority in the middle of an emergency, but it will pay off handsomely down the line.

The march of the coronavirus reminds us yet again that world leaders and philanthropic donors pay attention to epidemics only when an infection is on their doorsteps. They must recognize that the time to think about the next epidemic is now.

When it's fine to fail

The history of metrology holds valuable lessons for initiatives to reproduce results.

Everyone's talking about reproducibility – or at least they are in the biomedical and social sciences. The past decade has seen a growing recognition that results must be independently replicated before they can be accepted as true.

A focus on reproducibility is necessary in the physical sciences, too – an issue explored in this month's *Nature Physics*, in which two metrologists argue that reproducibility

should be viewed through a different lens. When results in the science of measurement cannot be reproduced, argue Martin Milton and Antonio Possolo, it's a sign of the scientific method at work – and an opportunity to promote public awareness of the research process (M. J. T. Milton and A. Possolo *Nature Phys.* **26**, 117–119; 2020).

The authors – at the International Bureau of Weights and Measures in Paris, and at the National Institutes of Standards and Technology in Gaithersburg, Maryland, respectively – draw on three case studies, each one an instalment in the quest to measure one of the fundamental constants of nature.

The researchers chose the speed of light (c); Planck's constant (h), a number that links the amount of energy a photon carries to its frequency; and the constant of gravitation (G), a measure of the strength of the gravitational force between two bodies.

For both Planck's constant and the speed of light, different laboratories have arrived at the same number using different methods – a sign of reproducibility. In the case of Planck's constant, there's now enough confidence in its value for it to become the basis of the International System of Units definition of the kilogram that was confirmed last May.

However, despite numerous experiments spanning three centuries, the precise value of G remains uncertain. The root of the uncertainty is not fully understood: it could be due to undiscovered errors in how the value is being measured; or it could indicate the need for new physics. One scenario being explored is that G could even vary over time, in which case scientists might have to revise their view that it has a fixed value.

If that were to happen – although physicists think it unlikely – it would be a good example of non-reproduced data being subjected to the scientific process: experimental results questioning a long-held theory, or pointing to the existence of another theory altogether.

Questions in biomedicine and in the social sciences do not reduce so cleanly to the determination of a fundamental constant of nature. Compared with metrology, experiments to reproduce results in fields such as cancer biology are likely to include many more sources of variability, which are fiendishly hard to control for.

But metrology reminds us that when researchers attempt to reproduce the results of experiments, they do so using a set of agreed – and highly precise – experimental standards, known in the measurement field as metrological traceability. It is this aspect, the authors contend, that helps to build trust and confidence in the research process.

One of the wider lessons from Milton and Possolo's commentary is that researchers from different domains must continue to talk and to share their experiences of reproducibility. At the same time, we should be careful about assuming that there's something inherently wrong when researchers cannot reproduce a result even when adhering to the best agreed standards.

Irreproducibility should not automatically be seen as a sign of failure. It can also be an indication that it's time to rethink our assumptions.

 Researchers from different domains must continue to talk and share their experiences."

Out-of-office should mean what it says

Employers must do more to support researchers when they take a break.

Setting an out-of-office e-mail reply should come with a sense of satisfaction. But in today's research world, an out-of-office message can seem little more than creative fiction. Its existence and the sender's absence will not bring work to a halt. They don't prevent an overworked researcher from feeling the need to check their inbox while away; nor do they stop senders attempting to contact people who are on holiday, and expecting a reply.

Some out-of-office messages do a better job. Last October, Stephana Cherak, an epidemiologist at the University of Calgary in Canada, received an impressive example from a colleague. "I do not respond to e-mails on weekends," it read. "If this is an emergency, please call my mobile. If you do not have my mobile number, then you do not have a weekend emergency."

Cherak approvingly tweeted the message. Of the more than 4,000 re-tweets and replies, many expressed support for drawing firm boundaries around time off, or offered their own tips. "My life has gotten much better since I decided that I don't need 'fastest/best/most consistent e-mail responder' to be part of my professional legacy," wrote @popmediaprof. And @runforbooze recommended that people politely write "I don't expect an immediate reply" if they have to send a message out of office hours.

We asked Cherak to reflect on this experience. In a column in *Nature's* Careers section, she had advice for all those trying to balance work with the rest of life (S. Cherak *Nature* **578**, 179–180; 2020). One recommendation is to ask for support from colleagues and supervisors.

Such support is vital, and employers must recognize that their staff need it. Indeed, in France, the 'right to disconnect' became law in 2017. Companies with more than 50 staff members are now obliged to discourage out-of-hours and holiday e-mail communication. Where changing the law isn't an option, a team of organizational psychologists at the University of Manchester, UK, has suggested setting up a 'bounce-back', so that e-mails received during time off are automatically returned to the sender.

There are several ways in which employers can support their staff when they take breaks, such as helping to put work on hold, accepting that projects will take a little longer and ensuring that essential tasks can be covered when colleagues are away.

Switching off from work is increasingly difficult – we at *Nature* struggle with this as much as does any organization. An out-of-office message must mean what it says if we are to have any hope of turning things around.

World view



By Nahid Bhadelia

Coronavirus: hospitals must learn from past pandemics

Use methods honed in previous outbreaks to prepare for the next one, says Nahid Bhadelia.

The world that is grappling with 2019 novel coronavirus (2019-nCoV) is different from how it was during the SARS and H1N1 pandemics. The disease itself, and information and disinfection, now travel faster than ever.

I worked as a clinician in West Africa during the Ebola outbreak, and in New York City hospitals during the H1N1 one. Now, I'm working in Boston, Massachusetts, to prepare for potential cases of 2019-nCoV acute respiratory disease. And many of the challenges are the same as those faced in previous outbreaks.

The specifics of each virus are important, but so is an overarching question: what do you do when large numbers of people arrive wanting care for suspected infections of an unfamiliar disease? This comes down to three decisions: how to quickly identify infected people, how to isolate and care for them and how to keep health-care workers safe.

As this epidemic grows, two trends will make it harder to identify people with 2019-nCoV infections while coping with those showing similar symptoms in the middle of the current influenza season. First, the 2013 and 2016 Ebola outbreaks taught us the importance of travel history. But with more countries reporting 2019-nCoV cases, it will be harder to teach hospital workers what locations to ask people about, and hospitals will need to devise strategies to keep staff aware of the changing geography of risk.

Second, as the H1N1 pandemic demonstrated, people with no relevant travel history will crowd emergency departments and other care settings. Hospitals and local public-health authorities will have to encourage people who are likely to be infected with 2019-nCoV to get diagnosed quickly while discouraging those infected with less-threatening diseases from seeking emergency treatment. Public-health authorities handle much of this education, but hospitals must strengthen communication among clinics and their patients.

Current data suggest that people could transmit the new disease before they show symptoms (C. Rothe *et al.* *N. Engl. J. Med.* <http://doi.org/gggjvr8>; 2020). Besides rapidly identifying travellers, hospitals must strengthen infection-control measures that apply to anyone with respiratory symptoms, such as by reinforcing hand hygiene and use of masks, frequently decontaminating crowded places and finding areas where patients with symptoms can be separated from others and cared for.

Most samples are still being shipped from hospitals to reference laboratories. A test closer to the bedside is crucial for quickly identifying people with 2019-nCoV



What do you do when large numbers of people arrive at facilities with an unfamiliar disease?"

and separating them from others with similar symptoms. Countries with confirmed cases are sharing viral genetic sequences – which makes developing tests easier.

Many hospitals in richer countries must decide whether patients should be cared for in specialized biocontainment units created for people with Ebola virus disease or in rooms assigned to those with other airborne diseases, such as tuberculosis and measles. But the demand for both could soon outstrip supply if the epidemic spreads, so hospitals could create a stepwise plan: one for dealing with a handful of patients, and another for when large numbers of sick patients cause a shortage of intensive-care beds. Hospitals might need to work with nearby facilities to ensure every person needing intensive care receives it.

Another dilemma hospitals face is deciding what personal protective equipment (PPE) health-care workers should use to keep themselves from getting infected. The Centers for Disease Control and World Health Organization advise that workers could prevent contact with body fluids, contaminated surfaces and virus particles in the air from sneezing and coughing using a range of ensembles: gloves and coveralls or gowns, paired with personal air-purifying respirators or certified particulate-filtering face masks.

Contrary to popular belief, the most protective option is not always the safest choice. Workers unaccustomed to complex PPE are more likely to use it incorrectly and thus put themselves at higher risk of infection. During the SARS epidemic, workers were at the highest risk of infection when putting on and taking off their PPE. Hospitals will need to continually train staff in using this equipment, and provide frequent re-enforcement. Also, restrictive PPE can affect the quality of care that patients receive. And uncommon PPE might be harder to get in large supplies. If supply changes mean that workers have to switch equipment mid-epidemic (as I experienced in West Africa), confusion soars. In the end, what works for each facility varies with resources and setting.

Hospitals will also have to manage illnesses among health-care staff. As more of their workers get sick, hospitals and clinics will have a harder time responding to the outbreak. But if health-care workers come in sick – and our experience in New York City during H1N1 showed that up to 60% of clinicians did so (N. Bhadelia *et al.* *Infect. Control Hosp. Epidemiol.* **34**, 825–831; 2013) – they could transmit the disease to patients and colleagues. Hospitals need staffing plans to cope with worker shortages.

Connecting these three sets of decisions is the fact that scientific knowledge about a disease changes and (ideally) increases as a new epidemic progresses. There is little guidance on how to craft policies and procedures while living through the uncertainty caused by a new virus. When this outbreak recedes, that guidance is where we must focus.

Nahid Bhadelia is an associate professor of infectious diseases and medical director of the special pathogens unit at the Boston University School of Medicine in Massachusetts.
e-mail: nbhadelia@bu.edu

News in brief

SUN'S ELUSIVE POLES TO BE IMAGED IN DETAIL FOR FIRST TIME

A European mission that will take the closest-ever pictures of the Sun and give scientists their first clear look at the star's uncharted poles launched on 9 February.

Equipped with 10 instruments, the €500-million (US\$550-million) Solar Orbiter, which took off from Cape Canaveral in Florida, will journey first to Mercury's orbit on a mission that could last 10 years.

"Nobody has been able to take images this close to the Sun before," says Helen O'Brien at Imperial College London, who manages the magnetometer instrument on the European Space Agency (ESA) mission, which also involves NASA. "We should see some beautiful images."

The mission's main aim is to investigate interactions between the Sun and its heliosphere – the bubble of the star's activity in space, says O'Brien. "It's really important to work out how the energy propagates from the surface out into interplanetary space."

The spacecraft (pictured on the left in this artist's

impression) will be placed into an orbit that will bring it, at its closest, just 42 million kilometres, or 0.28 astronomical units, from the Sun (1 AU is the distance between Earth and the Sun). It will take about two years to reach this orbit.

The Solar Orbiter's main science phase will begin in November 2021 and last for four years. But if the mission is extended, as ESA scientists hope it will be, the craft would enter a second phase, which would allow it to image the Sun's poles. Over several years, mission controllers would raise the angle of the spacecraft's orbit above the plane of the planets and up and over the Sun.

"That will give us the first-ever views of the solar poles," says Daniel Müller, a solar physicist at ESA's European Space Research and Technology Centre in Noordwijk, the Netherlands, who is the project scientist on the mission. "We believe that is key to better understanding the Sun's magnetic activity cycle." A previous mission, ESA and NASA's Ulysses spacecraft, flew over the poles in the 1990s and 2000s – but it had no cameras.



SCIENCE MINISTER'S CANCER CLAIMS SPARK CONTROVERSY

Excitement over the creation of Colombia's first Ministry of Science, Technology and Innovation has given way to anger and confusion over the appointment of Mabel Torres as science minister. The mycologist from the Technological University of El Chocó in Quibdó has made public claims about the cancer-fighting properties of a mushroom extract that she makes herself.

Torres says that she has given it to around 40 people with cancer – some of whom, she says, have entered into remission. But the treatment was not given under the auspices of a clinical trial, the methodology was not approved by a medical-ethics committee, and Torres has not submitted the results for publication in a peer-reviewed journal. Critics want her to resign; one fears that her appointment might embolden people peddling unproven medical treatments.

Torres defends her actions and says she has no plans to step down. "I haven't offered a drug, let alone marketed it. I have rigorously observed the established ethical protocols for scientific experimentation," she said in a statement.

Torres' supporters, who include prominent scientists, say she will be an advocate for marginalized regions – including El Chocó.



Pangolins suspected as source of coronavirus outbreak



Two researchers at the South China Agricultural University in Guangzhou have suggested that pangolins – long-snouted mammals often used in traditional Chinese medicine – are the probable animal source of the coronavirus outbreak causing global alarm.

Shen Yongyi and Xiao Lihua reported at a press conference on 7 February that they had identified the pangolin as the potential source of the coronavirus, on the basis of a genetic comparison of coronaviruses taken from the animals and from infected humans.

Scientists have already suggested that the virus originally came from bats, because of the similarity of its genetic sequence to those of other known coronaviruses, but the pathogen was probably transmitted to humans by another animal.

Researchers say the suggestion that pangolins spread the coronavirus to people seems plausible – but caution that the work is yet to be published.

The coronavirus has now infected tens of thousands of people globally, more than 1,000 of whom have died.

NASA SOARS WHILE OTHERS PLUMMET IN US BUDGET PROPOSAL

NASA could see a 12% increase to its US\$22.6-billion budget under a proposal released by US President Donald Trump on 10 February. As science agencies go, however, it is an outlier. The budget request, which covers all areas of government, cuts deeply across most research spending for the 2021 fiscal year, which begins on 1 October 2020.

The proposal includes \$38.7 billion for the US National Institutes of Health, about a 7% cut to current funding levels.

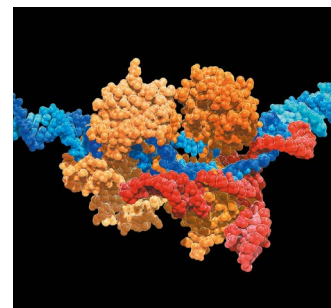
The US Department of Energy's Office of Science would lose nearly 17% from 2020 levels. It would also eliminate the popular Advanced Research Projects Agency–Energy, which received a record \$425 million last year, and slash the budget of the office of energy efficiency and renewable energy by 74%.

The proposal also seeks a \$500-million cut for the National Science Foundation. The agency's computer science and engineering section is the only one of its directorates that would see an increase, consistent with the administration's plans to prioritize artificial intelligence and quantum computing.

The Environmental Protection Agency's budget would be slashed by roughly 26%, to \$6.7 billion.

Although Congress has repeatedly rebuffed the president's requests for cuts to science – and has increased research spending – the budget proposal offers a view into the administration's priorities.

"Trump is being Trump," says Michael Lubell, a physicist at the City College of New York who tracks US science policy. "He can ask for what he wants, but it doesn't mean it's going to happen."



SUPER-PRECISE CRISPR TOOL ENHANCED BY ENZYME ENGINEERING

Researchers have boosted the accuracy of a technique based on the popular CRISPR–Cas9 genome-editing system by engineering enzymes that can precisely target DNA without introducing as many unwanted mutations.

The enzymes, reported on 10 February, could make a method called base editing more feasible as a tool to treat genetic diseases (J. L. Doman *et al. Nature Biotechnol.* [http://doi.org/dmgf](https://doi.org/dmgf); 2020).

Base editing uses the Cas9 enzyme to target DNA edits to a specific site, where other enzymes chemically convert one DNA base into another. This offers greater control than conventional CRISPR–Cas9 editing, but can still introduce 'off-target' changes at random locations in the genome.

A team led by David Liu, a chemical biologist at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, developed screening methods that can detect unwanted mutations without the need for costly full-genome sequencing. This allowed the team to identify new base-editing enzymes that can change the DNA base C to T without making as many off-target edits. The approach could allow researchers to develop safer gene therapies.

News in focus



BERNARD DUPONT (CC BY-SA 2.0)

A study on the social spider *Stegodyphus dumicola* was the first to be retracted.

'AVALANCHE' OF RETRACTIONS SHAKES BEHAVIOURAL-ECOLOGY COMMUNITY

Allegations of fabricated data in papers on spider behaviour have prompted a university investigation and some soul-searching.

By Giuliana Viglione

A complex web is unravelling in the field of spider research. On 5 February, McMaster University in Hamilton, Canada, confirmed that it was investigating allegations that behavioural ecologist Jonathan Pruitt had fabricated data in at least 17 papers that he had co-authored.

Since concerns about his work became public in late January, scientists have rushed to uncover the extent of questionable data in Pruitt's studies. Publishers are now trying to keep up with requests for retractions and investigations. So far, seven papers have

been retracted or are in the process of being retracted; five further retractions have been requested by Pruitt's co-authors; and researchers have flagged at least five more studies as containing possible data anomalies.

A tangled web

Pruitt, who is reportedly doing field research in Australia and the South Pacific, told *Science* last week that he had not fabricated or manipulated data in any way. He did not respond to multiple requests from *Nature* for comment on the mounting list of retractions, or the accusation that he had fabricated data.

His research looks at how different

personalities form in communities of social spider species that live in groups, and it has implications for emerging ideas on how animal behaviours evolve in the context of their environment.

The retractions started in mid-January, when authors of a paper in *The American Naturalist*¹ pulled it, citing "irregularities in the raw data". These were data that Pruitt had provided, showing how long it takes social spiders to resume typical behaviours after a disturbance, such as a simulated attack from a predator.

After a second retraction², Kate Laskowski, a behavioural ecologist at the University of

California, Davis, who had co-authored both studies with Pruitt, wrote a blogpost about those irregularities (see go.nature.com/39m535t). She had found multiple stretches of data that had been copied and pasted to represent findings for multiple spiders. When Pruitt's explanations failed to account for the anomalies, she requested that the journals retract the papers, reportedly with Pruitt's consent.

"Then, hell broke loose," says Niels Dingemanse, a behavioural ecologist at Ludwig Maximilian University in Munich, Germany, who has helped to uncover the data issues.

More than 20 scientists – co-authors, peers and other interested observers in the field – mobilized to pore through the data in almost 150 papers on which Pruitt is a co-author, looking for evidence of manipulated or fabricated numbers. They found similar signs of copy-and-paste duplications. In at least one instance, researchers identified formulae that had been inserted into a published Excel file, designed to add or subtract from a pasted value and create new data points.

Several have stated that they consider this clear evidence of fraud. Dingemanse says that his mind was made up by the "avalanche of retractions" in progress, as well as the mounting piles of irregular data. "It is hard to believe these data are not fabricated," he says.

The 17 papers that include questionable data have been cited more than 900 times, and it will take scientists a while to sort out which ideas have been supported elsewhere in the literature and which will need to be retested. "My guess is the impact will probably be pretty big," Laskowski says.

Pruitt had written "a lot of really impressive papers" and was regarded by many as a "rising star", says María Rebolledo-Gómez, a microbial ecologist at Yale University in New Haven, Connecticut.

A spokesperson for McMaster University confirmed that the institution was investigating, but would provide no further comment on issues of research integrity. The University of California, Santa Barbara, where Pruitt did most of the work in question, declined to comment on the specific case but said that it "would cooperate with any other institution conducting an investigation".

"My guess is that the impact will probably be pretty big."

Laskowski says that although the wave of retractions deals a blow to behavioural ecology, she is heartened by how quickly the community has acted to set the scientific record straight. Researchers have lessons to learn about making data publicly available – by one estimate, more than 60% of Pruitt's data-containing papers are in journals with no data-sharing requirements – and about checking data that they receive from colleagues. But she and others are optimistic that these lessons will ultimately strengthen the field.

1. Laskowski, K. L., Montiglio, P.-O. & Pruitt, J. N. *Am. Nat.* **187**, 776–785 (2016); retraction **195**, 393 (2020).
2. Laskowski, K. L. & Pruitt J. N. *Proc. R. Soc. B* **281**, 20133166 (2014); retraction **287**, 20200077 (2020).

suggested that they change the titles of their papers to mention an algorithm he had developed.

"The magnitude of his self-citation requests are shocking," says Jonathan Wren, an associate editor for *Bioinformatics*, a journal that last year barred Chou from reviewing its papers, although it did not name him at the time. "But what blows my mind is that suspicious citation patterns to him go back decades and authors comply with an apparently amazing frequency."

Chou retired from a career in the pharmaceutical industry in 2003. He then founded the Gordon Life Science Institute, which he calls an institute with "no physical boundaries", of which anyone can become a member. Before 2003, Chou had published 168 papers – mostly in the field of computational biology – which were cited around 2,000 times. But he now has 602 papers with more than 58,000 citations, according to Elsevier's Scopus citations database. He is one of the world's most highly cited researchers.

The *JTB* editorial says that Chou also handled papers written by close colleagues at his own institute – some of whom the journal later couldn't trace, which the editorial says calls into question their veracity. It adds that Chou sometimes reviewed papers under a pseudonym, or chose reviewers from his institution. And in many cases, Chou was added to papers as a co-author during the final stage of review.

"Regrettably, this process was repeated for dozens of papers," the editorial says. It adds that the journal wants to "apologize for missing this blatant misuse of the editorial system".

Chou told *Nature* that mentions of his algorithms in papers were "not from 'reviewer coercion', but from their very high efficacy and widely recognized by many users". But he declined to answer questions about the citation practices for which he was banned, and instead referred *Nature* to his website.

Wren flagged the suspicious citation patterns to the *JTB* after an investigation at his own journal. That probe revealed that in every review, Chou had requested that manuscript authors add citations – an average of 35 of them, 90% to papers he had co-authored. *Bioinformatics* announced that it had barred a referee in January 2019.

Wren, a bioinformatician at the Oklahoma Medical Research Foundation in Oklahoma City, says investigations into Chou's citations are under way at at least three other journals to which he has pointed out suspicious patterns. Wren is currently writing an algorithm to flag unusual citation patterns in papers automatically.

The case comes amid efforts by Elsevier to crack down on the practice of 'coercive citation'. Last year, the Amsterdam-based publisher said it was investigating hundreds of researchers whom it suspected of

JOURNAL BANS HIGHLY CITED RESEARCHER FOR CITATION ABUSE

Probe finds that Kuo-Chen Chou repeatedly suggested dozens of citations be added to papers.

By Richard Van Noorden

A US-based biophysicist who is one of the world's most highly cited researchers has been removed from the editorial board of one journal and barred as a reviewer for another, after repeatedly manipulating the peer-review process to amass citations to his own work.

On 29 January, three editors at the *Journal of Theoretical Biology* (*JTB*) announced in an editorial that the journal had investigated and

barred an unnamed editor from the board for "scientific misconduct of the highest order" (M. Chaplain *et al.* *J. Theoret. Biol.* **488**, 110171; 2020).

The journal's publisher, Elsevier, confirmed to *Nature* that the barred editor is Kuo-Chen Chou, who founded and runs an organization that he calls the Gordon Life Science Institute, in Boston, Massachusetts. According to the editorial, Chou asked authors of dozens of papers he was editing to cite a long list of his publications – sometimes more than 50 – and

manipulating peer review to boost their citations. Chou's case is the first to be revealed since that announcement. "While thankfully rare, such practices are an abuse of the peer-review system and undermine the hard work and commitment that editors and reviewers devote to ensuring the integrity of the scholarly record," a spokesperson says. "Elsevier has developed analytical tools to help detect such practices and is committed to implementing technology to flag citation manipulation before publication."

From 2014 to 2018, Chou was named as a highly cited researcher in a list produced by Clarivate Analytics, an information-services firm in Philadelphia, Pennsylvania, that owns the citation database Web of Science. But his name does not appear on the 2019 list; last year, Clarivate decided to remove scientists whose papers showed "unusually high levels of self-citation".

Elsevier hasn't yet decided what to do about papers that Chou handled that liberally cite his work, the spokesperson says.

microscopes and more sophisticated software for transforming the images they captured into sharper molecular structures. That paved the way for the current growth of cryo-EM, says Sjors Scheres, a structural biologist and specialist in the technique at the MRC Laboratory of Molecular Biology (LMB) in Cambridge, UK.

Richard Henderson, an LMB structural biologist who shared the 2017 Nobel Prize in Chemistry for his work developing the technique, says that even after these advances, growth was slow at first, because only a small number of labs had access to the equipment. But when they started using cryo-EM to produce detailed maps of molecules such as the ribosome – cells' protein-making machines – other scientists, as well as their institutions and funders, quickly took notice. "All the people who had invested in other things and made the wrong decisions, it took them a year to catch up," says Henderson.

He estimates that, by 2024, more protein structures will be determined by cryo-EM than by X-ray crystallography. Cryo-EM has already supplanted X-ray crystallography for one category of proteins that scientists are especially interested in – those embedded in cell membranes. Many such membrane-bound proteins are implicated in disease and serve as targets for drugs.

Advanced imaging

The structures of molecules determined by cryo-EM are also getting more detailed, thanks to continuing improvements in hardware and software, says Scheres.

Initially, the sharpest cryo-EM structures were of highly stable proteins that were used to test the limits of the technology. But Scheres has noticed that researchers are increasingly obtaining very high-resolution structures of medically important molecules, such as cell-membrane proteins, even though they tend to flop around.

"We're now coming to the point where the easy samples have been done and people are looking at more complex problems," says Ardan Patwardhan, a structural biologist at the European Molecular Biology Laboratory–European Bioinformatics Institute in Hinxton, UK, who leads the team that runs the EMDB.

Henderson expects the boom in cryo-EM structures to slow at some point. One factor that could sap growth, he says, is the high cost of the most powerful microscopes, which can exceed £5 million (US\$7 million). They also cost thousands of pounds each day to run, and require specialized labs that minimize vibrations. Henderson is campaigning to convince firms to develop cheaper, but still useful, microscopes that could spread the technique even further. "At the moment, you cannot go wrong by putting more investment into cryo-EM," he says.

THE PROTEIN-IMAGING TECHNIQUE TAKING OVER STRUCTURAL BIOLOGY

The number of structures being determined by cryo-electron microscopy is growing explosively.

By Ewen Callaway

A revolutionary technique for determining the 3D shape of proteins is booming. Last week, a database that collects protein and other molecular structures obtained using cryo-electron microscopy, or cryo-EM, acquired its 10,000th entry.

Submissions to the Electron Microscopy Data Bank (EMDB) – a popular repository for structures solved using electron microscopy – have increased exponentially in recent years, largely because of the explosive growth in the number of cryo-electron microscopes in laboratories worldwide (see 'Structure sleuths'). The EMDB curates structures solved with other microscopy methods, but the vast majority use cryo-EM.

The technique involves flash-freezing solutions of proteins or other biomolecules, and then bombarding them with electrons to produce microscope images of individual molecules. These are used to reconstruct the 3D shape, or structure, of the molecule. Such structures are useful for uncovering how proteins work, how they malfunction in disease and how to target them with drugs.

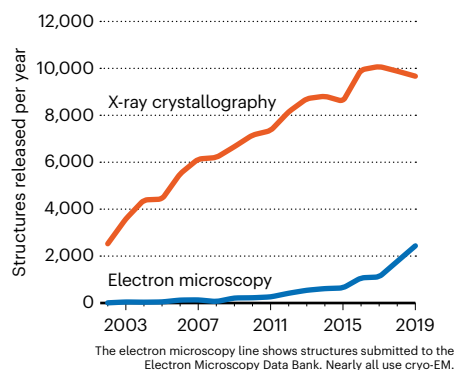
For decades, structural biologists preferred to use X-ray crystallography, a technique that involves crystallizing proteins, pummeling them with X-rays and reconstructing their shape from the resulting tell-tale patterns of diffracted light. X-ray crystallography produces high-quality structures, but it's not easy to use with all proteins – some can take months or years to crystallize, and others never crystallize at all. Cryo-EM doesn't require protein

crystals, but the technique languished because it tended to produce low-resolution structures – some scientists called it blobology.

Breakthroughs in hardware and software in 2012–13 produced more sensitive electron

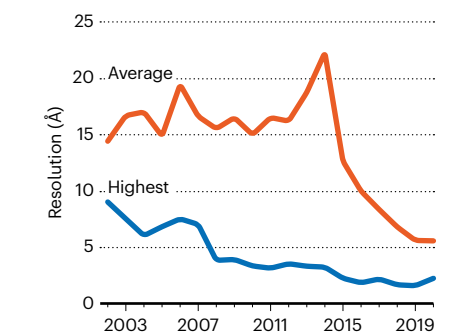
STRUCTURE SLEUTHS

Most structures of proteins and other biological molecules are still solved with X-ray crystallography. But a revolutionary technique called cryo-electron microscopy (cryo-EM) is catching up.



Fine detail

Cryo-EM can now resolve features that are less than 2 ångströms across.





BRETT GUNDLOCK FOR NATURE

A soldier patrols La Pampa, an area in the Peruvian Amazon that was once lush rainforest.

CAN A RAINFOREST DESTROYED BY GOLD-MINERS BOUNCE BACK?

Illegal miners have left La Pampa, giving researchers access to an inadvertent experiment in restoration.

By Jeff Tollefson
in La Pampa, Peru

Holy shit!” Miles Silman gasped as his motorized rickshaw rattled out of the forest and onto a desolate beach. All traces of the trees, vines and swamps that once covered this patch of the Amazon rainforest had vanished. In their place were sun-baked dunes and polluted ponds created by illegal gold-mining. Silman, a conservation biologist at Wake Forest University in Winston-Salem, North Carolina, was there to document the carnage.

La Pampa was once the largest and most dangerous gold-mining zone in the Peruvian Amazon. It was so riddled with gangsters that scientists dared not enter, and, for nearly a

decade, could only watch by satellite as gold hunters mowed down some of the most biodiverse rainforest on the planet. That ended in February 2019, when the government declared martial law and expelled an estimated 5,000 miners.

Now, La Pampa is deserted and under military guard. When Silman and his colleagues surveyed the area for the first time in late June, they found a barren, eerily quiet landscape polluted with mercury, a toxic by-product of mining. The data that the researchers collect during this inadvertent experiment could help to determine the extent to which restoration is possible – or document the evolution of an entirely new, and human-made, ecosystem.

Silman and his colleagues at the Center for

Amazonian Science and Innovation (CINCIA), a non-profit research institute in Puerto Maldonado, Peru, have spent the past several months mapping the area with drones and surveying the remaining plants and animals. The team has been studying dozens of tree species to see which can survive among the dunes and along the shores of ponds.

CINCIA scientists have also tested the air, water and soil for mercury contamination. Another team, from Duke University in Durham, North Carolina, has collected data there to help unravel how mercury – which can harm children’s brain development – moves from polluted water or soil and up the food chain.

The research by the CINCIA team and other scientists will feed into the Peruvian government’s ongoing efforts to rehabilitate the area, says Camila Alva, director of pollution control and chemical substances at the country’s environment ministry.

The government has already begun a pilot project to restore the Tambopata National Reserve, a protected forest that miners invaded when La Pampa expanded. Peruvian President Martín Vizcarra visited the reserve on 5 December in a show of support. Results from that work could help to guide the government’s longer-term efforts to reforest, and perhaps even resettle, parts of La Pampa.

La Pampa started out as a roadside outpost

This story was supported by the Pulitzer Center.

on the Interoceanic Highway, one of the first stops for would-be miners who flowed from the Andes into the Madre de Dios region of the Peruvian Amazon when gold prices spiked a decade ago. The regional capital, Puerto Maldonado, grew into a mining hub. And La Pampa became a bustling town of some 25,000 people, with a reputation for prostitution, modern slavery and organized crime.

The boom proved too lucrative to control. The gold dust is almost everywhere, and with a few rudimentary devices – including a petrol-powered water pump and a hand-made sluice – anyone can collect silt. Then it's just a matter of mixing in mercury, which binds the gold, to recover as much as 10–15 grams of gold per day. That is several hundred dollars' worth on the global market.

Miners don't worry about mercury's inevitable release into the environment or the health effects of exposure, says Luis Fernandez, CINCIA's executive director.

But researchers want to understand how much mercury miners left behind, and how it is moving through the ecosystem. On a sunny June day in La Pampa, the CINCIA team explored the site in preparation for research that will look for mercury contamination in the air, water and soil, as well as birds, fish and other aquatic life.

"We have birds and insects, that's something that we can sample," said environmental chemist Claudia Vega, who coordinates CINCIA's mercury research programme. The tests will help to determine how much mercury could be moving into the food chain, where it would pose a danger to people, including the farmers who have laid down claims for land in La Pampa.

So far, testing by CINCIA and other researchers suggests that the mercury contamination is concentrated in the ponds. That means the land is probably safe for farming, but that eating fish that live in the ponds could be dangerous. "We cannot put people out there – families and children – if we don't know what the risk is," says Martin Arana, a forest engineer who is advising the Peruvian forest service.

Drone maps dunes

The team is also measuring the extent of deforestation in La Pampa and the potential to reforest the area. During their trip in June, the researchers launched their first drone flight on a dune deep inside the mining zone. The drone flew north–south transects at a height of around 200 metres to produce detailed 3D maps of the area's topography, including its dunes and ponds.

Silman's team is using those maps to estimate how much carbon was released into the atmosphere when the forest was mowed down to make way for mining activities. That information can also be used to track forest recovery and guide future plant surveys.



Illegal gold-mining has transformed forested land into sand and ponds.

Government officials are assessing the cost and technical feasibility of a major reforestation effort – as well as the jobs that it might produce. Working with CINCIA, Peru's park service and environment ministry have already launched their pilot reforestation project on 30 hectares of the Tambopata National Reserve. The agencies are planning to replicate that work across more than 750 hectares in the reserve.

The forest service is also studying how to design, implement and pay for an even larger reforestation project in La Pampa that could begin in a few years. But Arana says that the government will have to remain vigilant to

"It's going to be a hell of a lot better than a barren landscape with some toxic puddles in the middle."

the threat of illegal mining. "What happens if the price of gold is very, very high?" he asks. "Maybe the illegal miners come back to La Pampa, and there will be conflict with the people who are working in reforestation."

Silman is interested in understanding how different types of vegetation will recolonize the landscape naturally – and whether people might be able to guide and accelerate the process of reforestation. Strangler fig trees, which typically start life high in tree tops where light is abundant and then strangle their hosts, are already sprouting up in La Pampa's dunes, alongside burrowing owls (*Athene cunicularia*) that typically nest in arid shrublands.

Silman's team has been growing test plots of more than 75 plant species to guide the reforestation push. The scientists are tracking how the plants perform in a variety of

conditions: some prefer flat terrain with direct sun, whereas others need shade or very moist soil. The team's results suggest that adding charcoal – or a similar substance called biochar – to the soil bolsters plant growth and survival (D. Lefebvre *et al. Forests* **10**, 678; 2019). "We want to give people options, so that we aren't just planting trees that are going to die," Silman says.

Miners stripped the region's soils of all of their nutrients and fundamentally changed the way water moves through the landscape. That will make restoration attempts in La Pampa more difficult than in other areas where people have rebuilt or restored ecosystems harmed by mining, says Stuart Pimm, an ecologist at Duke University.

But rather than worrying too much about trying to recreate what was there before, Pimm says that scientists and the government should get some plants in the ground and let nature take its course. "Just getting some forest cover is something they can probably do," he says, "and it's going to be a hell of a lot better than a barren landscape with some toxic puddles in the middle."

As Silman and his colleagues wrapped up their day of field work in June, the Sun was setting – and La Pampa was coming alive. Ducks were on the move, and fish in ponds began rising to feed on insects.

Silman has little doubt that plants and animals will recolonize this largely empty space over hundreds or thousands of years. The question, he says, is whether scientists can help to accelerate that recovery, or whether La Pampa will remain little more than a monument to human stupidity over the coming decades.

"That land has already been deforested," he says. "There's a lot of incentive for us to be clever and to try to do good things there."



BRINGING BACK THE BOGS

Around the globe, drained peatlands are emitting billions of tonnes of carbon dioxide each year. To keep climate change in check, governments and researchers are working to keep peatlands healthy. **By Virginia Gewin**



A flux tower in Scotland's Flow Country measures gas concentrations and other variables in a peatland.

they lock up vast amounts of carbon. If they are not kept healthy, the bogs could release their stored carbon and accelerate global warming.

That's why a team of researchers and land managers is digging up trees and flattening furrows in former plantations southwest of Thurso. The effort is part of a roughly £50-million (US\$65-million) investment that the Scottish government and other organizations have made towards restoring the country's blanket bogs – undulating carpets of spongy hummocks built from *Sphagnum* mosses. The largest area of blanket bogs in the world is in the Flow Country – a low-lying expanse between sheer cliffs to the north and glacially carved mountains to the southwest.

Remote and exposed, these peatlands are named after the Norse term *floi*, which means boggy ground. They have long been described as worthless wastelands. “Local people called the peatlands *mamba* – miles and miles of bugger all,” says Roxane Andersen, a biogeochemist at the University of the Highlands and Islands' Environmental Research Institute in Thurso.

More than 80% of the 1.7 million hectares of peatland in Scotland have been cut for fuel or otherwise degraded, and roughly 500,000 hectares have been drained and forested with non-native conifers. “The reality, though, is the trees did poorly,” says Andersen.

Despite that, the peatlands have tremendous value for carbon storage. These areas hold more than one-quarter of all soil carbon, even though they account for only 3% of Earth's land area¹. Globally, peatlands hold more than twice as much carbon as the world's forests do, according to the United Nations Environment Programme.

But in many places, humans have turned vast expanses of these environments from long-term carbon sinks into carbon sources. Damaged or drained peatlands worldwide emit at least 2 billion tonnes of carbon dioxide annually – roughly 5% of anthropogenic greenhouse-gas emissions – largely through peat fires and oxidation of the buried carbon. And emissions from bogs are expected to rise sharply.

As the threat of climate change has grown more severe, researchers and governments have identified peatlands as ideal targets for stopping emissions, and even sopping up carbon. Although Canada, Russia and Indonesia contain the largest tracts of peatland in the world, Scotland has emerged as a leader in the effort to restore the habitat, which covers more than 20% of the country (see ‘For peat's sake’). Scotland will probably meet, if not exceed, its 2020 goal of restoring 50,000

hectares, mainly on government-owned nature reserves and forestry land. And it aims to push that total to 250,000 hectares by 2030.

Restoring peatlands to health is one of the key ways in which Scotland, which last April became the first country to declare a climate emergency, intends to reach net-zero greenhouse-gas emissions by 2045. “Scotland has raced out in front by making good connections with researchers and government,” says Jack Rieley, a tropical-peatland ecologist and executive board member of the International Peatland Society, which is based in Jyväskylä, Finland. Researchers from around the world have flocked to Scotland to glean insights into how to develop a successful national strategy for restoring peatland.

The biggest question is whether restoration will simply stop carbon emissions from peatlands or revive the bogs to the point that they can store more carbon. Other countries, notably Indonesia, are also pursuing efforts to reduce carbon losses from their peatlands. To make sure that these projects are working, researchers are developing satellite techniques and other tools to monitor the health of these landscapes.

But there is no guarantee that the efforts will pay off. “It's so easy to break an ecosystem, and it's so hard to bring it back,” says Andersen. “We can't recreate something from the past, but we can do our best to make it resilient.”

Tough going

Just over 100 kilometres southwest of Thurso, the boggy soil is so sodden in spots that I sink up to my knees and nearly lose a boot. But the muck hasn't stopped two excavators – each more than 13 tonnes – that are fitted with

“It's so easy to break an ecosystem, and it's so hard to bring it back.”

extra-wide tracks to distribute their weight. As part of an effort to convert the region back to bogs, they trundle across the peat, cutting and stacking stands of trees that have been there for 30 years.

The timber is low quality, pockmarked by hungry pests and prone to being blown down, a hallmark of trees that are growing in acidic peat. Neil McInnes and Tim Cockerill oversee this and other restoration projects undertaken by Forestry and Land Scotland, a government land-management agency based in Inverness. The harvest costs more than the timber is worth, and because the trees will be either incinerated on site to generate electricity or made into heating pellets, the carbon in the trees will return to the atmosphere.

Removing the trees was a bitter pill at first. Many foresters felt they were being unfairly

On a chilly September morning in Scotland's northern highlands, a giant excavator rumbles back and forth across peatlands that stretch to the horizon. As the wind whips across the mossy terrain, the machine's operator is undoing decades of damage by smoothing out the drainage ditches that scar the landscape.

The peat here can reach up to 10 metres deep and developed slowly over thousands of years. Then, in the middle of the twentieth century, Scotland embarked on an ill-fated effort to transform the bogs into tree farms. Land-owners ploughed trenches to drain bogs and planted pine trees and spruce that often failed to thrive. As the ventures struggled, researchers and the Scottish government started to see the peatlands in a fresh light, recognizing that

Feature

criticized for having planted them in the first place – even though it had been a government directive at the time. But McInnes says that attitudes have changed over the past few years as people have grown to understand the carbon-storage potential of peatlands, and the Scottish government has made it a priority to reduce emissions. “It doesn’t feel like a fight any more,” he says.

Early peatland-restoration efforts began in Flow Country in 1995, focused more on restoring bird habitats. “Carbon was barely on the agenda at that time,” says Norrie Russell, former manager of the Forsinard Flows reserve, which is owned by the Royal Society for the Protection of Birds and is where Andersen conducts her research.

The agenda gained momentum in 2010, when the International Union for Conservation of Nature launched the UK Commission of Inquiry on Peatlands to assess the state of these ecosystems. That effort – along with widespread support for tackling climate change – triggered more interest in nursing peatlands back to health. Now, Russell says, the political push for peatland restoration is focused mainly on keeping carbon locked up. In a 2017 public survey (see go.nature.com/2s-fvbiy), the vast majority of respondents supported restoration to mitigate climate change, to improve water quality and wildlife habitat and to protect this important aspect of Scotland’s identity.

Towers of resilience

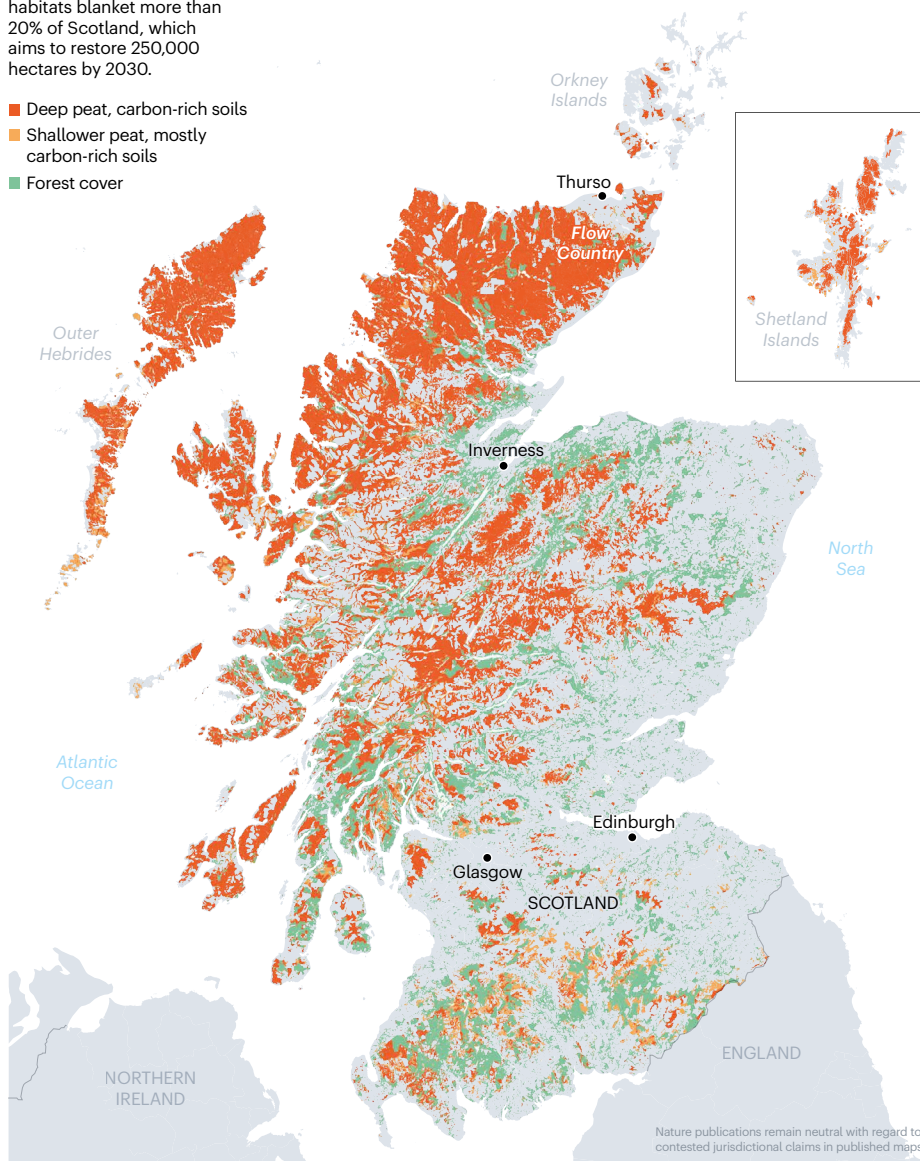
Andersen is working with McInnes and Cockerrill, as well as various organizations, to determine how best to manage the land for carbon storage. To gather evidence, she and her colleagues have installed four towers in Flow Country since 2008 to monitor the flow of gases and temperature, among other variables. Sensors near the towers measure heat flux, water level, soil temperature and precipitation. Building on existing data, Andersen won a £986,088 award last year from the London-based charity the Leverhulme Trust to determine how to make peatland resilient.

In the data collected so far, Andersen and her colleagues have detected some promising changes². They found that the first patches of restored peatlands, in which trees were simply cut and rolled into the blocked drainage ditches, switched from a carbon source to a carbon sink after 16 years. Although that work demonstrated that transitioning forest back to bog can be an effective way to restore a carbon sink, the researchers found that they could get faster results with more intensive management – such as clearing the carbon-rich trees and branches and flattening the ground. Although these more intensive strategies can trigger an initial pulse of greenhouse-gas emissions by disturbing the soil, once it is more uniformly wet this can also accelerate the switch

FOR PEAT’S SAKE

Peatlands hold more than one-quarter of the planet’s soil carbon, yet cover just 3% of the land. These habitats blanket more than 20% of Scotland, which aims to restore 250,000 hectares by 2030.

- Deep peat, carbon-rich soils
- Shallower peat, mostly carbon-rich soils
- Forest cover



from carbon source to sink – bringing it down to as little as ten years, says Andersen.

These results mirror research in Canada that found it takes one to two decades for peatlands to recover following restoration efforts³. The trick to restoring the natural hydrology, the way water moves through the system and is stored by the peat, is choosing locations that aren’t too degraded and where there is still enough residual peat and plant vegetation, says Nigel Roulet, a peatland scientist at McGill University in Montreal, Canada. “If you nudge systems along, and pamper them through first

years of recovery, they take off on their own,” says Roulet, “and carbon dynamics return to a natural system within a decade or two.”

But that’s a complicated story to convey – especially amid a groundswell of support around the globe for efforts to plant trees to combat global warming. Last year, a study suggested that Earth’s ecosystems could support 1 billion more hectares of forest – and store 25% of the atmospheric carbon pool⁴. Politicians in many countries, including the United Kingdom, have been eagerly promoting efforts to plant more trees. Scotland planted

SOURCES: Z. YU ET AL. *GEOPHYS. RES. LETT.* 37, 113402 (2010); SCOTTISH NATURAL HERITAGE

11,200 hectares of new woodlands in 2018. And in the run-up to last December's UK general election, both the Labour and Conservative parties promised to plant millions more trees each year. These new arboreal ambitions could make it harder for researchers and officials to argue that peatlands are the wrong places for trees. "Unless landowners and managers all work together on an agreed strategy then there will be pressure," says McInnes. "We've seen this before."

Breathing bogs

The key question about restoration efforts across the globe is how well they can slow greenhouse-gas emissions from bogs. To answer that, researchers need cheaper and faster tools for assessing the health of peatland over wide areas. Andersen has partnered with geoscientist David Large at the University of Nottingham, UK, to develop a method for

monitoring 'bog breathing' through satellite measurements – specifically, interferometric synthetic aperture radar (InSAR). Because peatlands that are functioning well rise and fall with the level of the water table, the carbon emissions can be inferred from how the peat behaves, says Large.

The team tested this method on 22 sites around the Flow Country over 18 months and found that wet, mossy peat in good condition – the least likely to be a carbon source – rises in mid-winter and falls in mid-summer⁵. Drier, shrubby peat, which is more likely to emit carbon, rises in late spring and falls in late summer. As a next step, the researchers plan to correlate their InSAR results with measurements of carbon emissions.

InSAR will offer funders and government officials a means of quantifying success, says Large. "At what point is peat restored? We've spent millions and haven't really thought

through what success will look like," he says, at least in terms of metrics. Large is now testing the tool in tropical peatlands, which he says are challenging because in areas such as southeast Asia, peat builds only under forest cover, and the trees cause trouble for InSAR. If the methodology can be validated across peatland types and conditions, it could help governments to choose which areas to restore and to monitor how effective interventions have been, says Susan Page at the University of Leicester, UK, who studies peatlands in southeast Asia.

Other teams are developing different methods for monitoring peatland emissions. In the tropics, for example, researchers are tracking deforestation, which often precedes efforts to drain the peatlands. Every country will have to develop its own monitoring system, says Hans Joosten, a peatland ecologist at the University of Greifswald in Germany.

Monitoring is urgently needed in many regions, including Indonesia. The country is plagued by seasonal fires that spread over dry peatlands and send billows of smoke across much of the country. The fire risk has increased in the past few decades because dams were installed to drain the country's peat and grow crops – notably oil palm trees, which do best when the water table is roughly 80 centimetres below the surface. Following devastating peat fires in 2015, Indonesia set an ambitious goal to restore 2 million hectares, about 10% of the roughly 20 million hectares of the country's original peat swamp forests, by 2020 to prevent fires and improve air quality.

By the end of last year, the campaign has re-wetted about 788,000 hectares, which involves raising the water table to within 40 centimetres of the surface. Nazir Foead, head of Indonesia's Peatland Restoration Agency says investigations in the country found that "when the table fell below 40 centimetres, the fire incidences soar significantly". Indonesia plans to achieve more than half of its carbon-reduction goals to support the Paris climate agreement through re-wetting and protecting peatlands.

In theory, these plans should reduce Indonesia's emissions, but they probably won't restore peat's ability to store new carbon, according to several researchers. "Re-wetting is the initial stage towards peatland restoration but it is not the magic bullet," says Rieley. Unlike in Scotland where mosses build up peat, trees are needed to deposit peat layers in tropical systems. In Indonesia, "where is the peat going to come from?" asks Rieley. Foead says his agency can't yet quantify how many trees have actually been replanted.

Even if Indonesia doesn't turn its peatlands back into a carbon sink, Joosten argues that re-wetting to 40 centimetres below the peat surface will reap big rewards from a climate perspective. Doing so would cut emissions from re-wetted areas by 50% because it halves



Scotland's Flow Country is the world's largest area of blanket bogs.

the amount of peat exposed to oxidizing conditions. And it would reduce global emissions much more than Scotland's endeavours, says Joosten, who was part of an international team that, in 2018, won the Indonesia Peat Prize, which is awarded by the government and the David and Lucile Packard Foundation, based in Los Altos, California. The team devised a method to map the extent and depth of peat.

Rare efforts

A fundamental problem is that large-scale peatland restoration is happening in just a few locations, say researchers. In fact, the global total peatland area is decreasing because bogs continue to be drained in the tropics and the land is converted for other uses. If that continues, carbon released from peatlands will help to send the global temperature shooting past the target of 1.5–2 °C warming above pre-industrial levels set by the Paris agreement.

One complication in the effort to re-wet peatlands is that restored wetlands will produce some amount of methane, which is a potent greenhouse gas. But Joosten says that this will be more than balanced by the reduction in emissions of carbon dioxide and nitrous oxide. Overall, re-wetting has a net benefit for the climate. Rather than aiming to turn global peatlands into sinks, he says, a more realistic near-term goal is to make bogs carbon neutral.

Achieving carbon neutrality for peatlands across the globe would have a major impact. Last year, Page and her colleagues found that by 2015, drained peatlands had emitted about 80 billion tonnes of carbon dioxide – and that this cumulative amount would roughly triple by 2100 (ref. 6). Estimates suggest that nations will need to limit future carbon-dioxide emissions to something on the order of 400 billion to 1,600 billion tonnes to keep temperatures from rising above the Paris target. But peatlands are on track to account for roughly 10–40% of that budget, unless countries take steps to protect and restore these environments, according to Page and her colleagues.

To keep that from happening, says Joosten, “all drained peatlands in the world have to be re-wetted. No cherry-picking which are easiest, cheapest or most effective any more”. Indeed, the United Nations Environment Assembly adopted its first ever peatland resolution last year, urging member states to conserve and restore these carbon-rich ecosystems.

Still, researchers say it will be important to document how much carbon is lost or stored in different peatlands, so that countries can meet their targets for the Paris climate accord and future agreements. And basic information about peatlands – including their extent and depth – is still lacking in many areas. Just three years ago, scientists discovered the world's largest continuous tropical peatland in the Congo basin of central Africa⁷.

“It is impossible to monitor greenhouse-gas

emissions over such large areas directly in practice – no country in the world does that,” says Joosten. In Indonesia, non-governmental organizations have highlighted that there is no independent monitoring of re-wetting effectiveness, he says.

And despite efforts to raise the water table in large swathes of Indonesia's peatlands, the country faced one of its worst fire seasons in 2019. “The areas that burnt were sites that were restored,” says Lahiru Wijedasa, a peatland ecologist at the National University of Singapore who is studying Indonesia's peatlands.

“At what point is peat restored? We've spent millions and haven't really thought through what success will look like.”

“We are at the early stages of understanding how these ecosystems function as a whole,” he says. The fires call into question whether Indonesia's degraded peatlands can be restored and how they will respond in the future, says Wijedasa.

Andersen agrees. “If degradation is too extensive, are we at risk of losing peatland areas before we can do anything about it?”

Burning questions

On 12 May 2019, a fire broke out on one of Andersen's restoration sites in Scotland. She recalls sleepless nights spent tracking the fast-moving blaze as it burnt more than 50 square kilometres. “It looked apocalyptic with an orange sky and dark clouds of smoke,” she says. “You could hardly breathe or see.” But what was most impressive, she recounts, is the speed at which it travelled. “It basically covered nearly 15 kilometres in one day.”

Andersen says that unusually hot, dry conditions preceded the fire and left the *Sphagnum* moss brittle. “The rivers were the lowest they've been since 1976.” Serendipitously, a couple of the driest sites were part of the InSAR validation study. The researchers found that the surface of the peat that had been most affected by the drought had collapsed, and it hadn't recovered when it began to rain again before the fire. “We saw consequences that outlast the drought for a long period of time,” she says.

Still, the restoration efforts seemed to help. Areas that had good *Sphagnum* cover and remained wet despite the drought had only low or medium fire damage, compared with spots that were still actively drained and had only patchy *Sphagnum* cover, which received the deepest burns and damage, according to Andersen.

Three weeks after the fire, she and her colleagues submitted a successful grant proposal to the UK Natural Environment Research

Council to study the impact of the blazes. The team will use ground measures, images from crewless aerial vehicles, and InSAR data to compare different types of peatland management – some had been restored more intensively, whereas others had been left to recover with fewer interventions. The researchers will assess how severely the peat burnt in each area, how it recovers and how much carbon was lost. They have also installed a fifth flux tower in the burnt area to measure how the fires affect carbon emissions. These data will be useful as researchers determine how best to restore sites to withstand future climate stresses, says Andersen.

Scotland has several advantages over other regions in its quest to restore peatlands – for example, landowners in the sparsely populated Flow Country can still make a living from restored peatlands, typically through tourism related to hunting and fishing. In Indonesia, however, people struggle to find crops that will grow on wet peaty soils and provide livelihoods for residents.

I was able to see at first hand some of the impacts of Scotland's efforts last year during a slog through the rain at the Langwell and Braemore estate. Roughly 6,000 newly installed dams have stymied erosion on grounds used for stag hunting and fishing. Between the dams, the water has pooled and is dotted with iridescent mosses. Anson MacAuslan was among the first estate managers to secure funding from Peatland Action – a project funded by the Scottish government to restore peatlands. He has spent roughly £185,000 on restoring 7% of the 19,000-hectare estate. He has already seen direct benefits from the dams, which have reduced flooding risk and improved water quality in the streams where salmon swim.

As several of the neighbouring estates start their own restoration projects, Andersen says that the shift in public perception of peatlands has been a key legacy of the Flow Country restoration project. There is even an effort afoot to nominate the Flow Country as a United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage site – which would be a first for a peatland. Although people used to call this landscape worthless, she says, “we don't hear that any more”.

Virginia Gewin, a science journalist in Portland, Oregon, reported this story with support from the European Geosciences Union.

1. Turetsky, M. R. et al. *Nature Geosci.* **8**, 11–14 (2015).
2. Hambley, G. et al. *Mires Peat* **23**, 5 (2019).
3. Nugent, K. A. et al. *Environ. Res. Lett.* **14**, 124030 (2019).
4. Bastin, J.-F. et al. *Science* **365**, 76–79 (2019).
5. Alshammari, L. et al. *JGR Biogeosciences* <https://doi.org/10.1029/2018JG004953> (2019).
6. Leifeld, J., Wüst-Galley, C. & Page, S. *Nature Clim. Change* **9**, 945–947 (2019).
7. Dargie, G. C. et al. *Nature* **542**, 86–90 (2017).

Books & arts



M. LIVIO AND THE HUBBLE TEAM (STSCI/NASA/ESA)

A cloud of interstellar gas and dust, captured by NASA's Hubble Space Telescope.

From Big Bang to cosmic bounce

A physicist and humanist takes us on a grand tour of all time. **By Philip Ball**

Brian Greene's *Until the End of Time* sits within a tradition of grand, synoptic visions of the Universe, rooted in physics, that feels (to this British reader) distinctively American. Halfway through, I realized why. With its scepticism of religion but openness to humanistic wonder, awe of nature, celebration of the individual and recognition of the power of physical law, the narrative has a strong whiff of transcendentalism. There is an echo of philosopher Henry David Thoreau in Greene's account of lying out at

night, enraptured by the aurora borealis. And essayist Ralph Waldo Emerson's declaration that the "sublime laws play indifferently through atoms and galaxies" could almost be this book's epigraph.

Such qualities lift this work above many accounts of the cosmic story spanning from the Big Bang to the end of time – whether that's a big rip, heat death or cosmic bounce. Greene takes us from quarks to consciousness, and from the origin of life to the genesis of language. He draws from an impressive range of sources, such as poets William Butler Yeats and Sylvia Plath. In attempting to weave in the evolution of physical laws with that of the human mind and cultures, Greene's aim vaults beyond that of his bestselling 1999 book, *The Elegant Universe. Until the End of Time* is packed with ideas; whether they come together as a convincing story is another matter.

This narrative features humanity as a brief moment when matter became self-aware. Current physical and cosmological theories imply that this state of affairs can't last. Eventually proton decay, a dominance of dark energy or thermodynamic heat death will doom all matter and thought. Greene, however, suggests that intelligent beings could eke out their thought processes almost indefinitely by gradually slowing them to minimize their inevitable thermodynamic cost.

He views this extinction of sentience as a cosmic tragedy. It's poignant to see a modern physicist, however girded with string theory, the general theory of relativity and the equations of quantum mechanics, experience the same anguish that goaded ancient monarchs to defy mortality by commissioning monumental tombs. Greene finds the solace that religion typically provides in the idea that the "small collection of the universe's particles" that constitutes humanity can evolve and "with a flitting burst of activity create beauty, establish connection, and illuminate mystery".

His grand tour is sometimes breathtaking, necessarily selective and occasionally superficial. It often lacks the space or rigour to do its vast range of subjects justice. Beyond fundamental physics, Greene is a lucid summarizer of other popular accounts, but little more. That can leave his story patchy, and even misleading at times. His explanation for why water is a special solvent required for life attributes it all to the molecule's polar nature – in which case it would not be special at all. (Hydrogen bonding is left out,

and although that does not tell the whole story, neglecting it means we get almost no story at all.) To explain the origin of myths, the book offers a bit of obsolete early-twentieth-century anthropology from the likes of folklorist James George Frazer, that is given a contemporary gloss of evolutionary psychology.

The biggest shortfall is in the account of how biology works, which seems to be derived largely from physicist Erwin Schrödinger's 1944 book *What Is Life?* and biologist Richard Dawkins's 1976 *The*

"Greene remains wedded to the idea that the most reductive view has ultimate authority."

Selfish Gene. Life in Greene's reckoning is all encoded in the genome, and once molecular replicators appeared on the planet, the rest was just evolutionary history. He adds that non-equilibrium thermodynamics can give us a head start: its tendency to create spontaneous knots and patterns of local order are a stepping stone towards life's organization. But what's missing – foreshadowing a wider lacuna in the book – is any sense that intermediate levels of that organization, particularly the cell, are equally fundamental.

When it comes to human behaviour – creativity, art, story, religion – Greene places a reductive faith in evolutionary psychology. He is probably right to say that many of our complex behaviours are underpinned by rather basic adaptive impulses, but he doesn't adequately acknowledge how culture shapes them. For instance, he supports psychologist Steven Pinker's notorious description of music as "auditory cheesecake". This posits that music is enjoyable because it piggybacks on capacities that evolved for other reasons, such as the ability to separate our auditory experience into comprehensible chunks. This might or might not be true, but

to appreciate what music really means, we need to consider its cultural, historical and social specifics, and not just attribute it to "our ancient adaptive sensitivity to sounds with elevated information content".

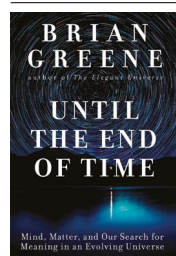
Whether in cell biology or a musical tradition, asking why any specific feature is the way it is demands that we consider a causal explanation. And therein lies the problem with Greene's approach: where it seeks out cause.

It's true that when he enlists physics as the underpinning theory of everything ("Life is physics orchestrated"), this is not the physicist's standard hubristic claim. Indeed, he points out that we need "overlapping narratives" for explanations of phenomena at different scales of size and complexity, from subatomic particles to galaxies, each of which must at least be consistent with the one below. And Greene acknowledges that an account of human behaviour at the level of fundamental particles would be pointless. But he still implies that causation flows upwards through the hierarchy of scales. We lack genuine free will, he says, because there is no such factor at play among the fundamental forces.

Thus, Greene remains wedded to the idea that the most reductive view has ultimate authority – that it all comes down to particles, entropy and evolution. "Perhaps one day we will invoke a unified theory of particulate ingredients to explain the overwhelming vision of a Rodin," he writes. He doesn't recognize that in complex systems, new properties and causative mechanisms that arise at only the higher levels of the hierarchy are as real and fundamental as, say, the strong and weak nuclear forces. This is what physics Nobel laureate Philip Anderson argued in his 1972 essay 'More Is Different'.

If we accept Anderson's position, we have to call into question the entire programme that Greene articulates here. By the time we get to, say, the human impulse to create stories, are Big Bang cosmology and quantum mechanics meaningful parts of the narrative? Perhaps, then, by setting out a vision of the world as seen by a thoughtful, humanistic fundamental physicist, Greene has offered not so much a state-of-play panorama as a tour showing where that view works spectacularly and where it falls short. It is an eloquent invitation to debate.

Philip Ball is a science writer and author; his latest book is *How To Grow a Human*. e-mail: p.ball@btinternet.com



Until the End of Time: Mind, Matter, and Our Search for Meaning in an Evolving Universe
Brian Greene
Penguin (2020)

Jack Baldwin

(1938–2020)

Organic chemist whose rules aided the synthesis of natural products.

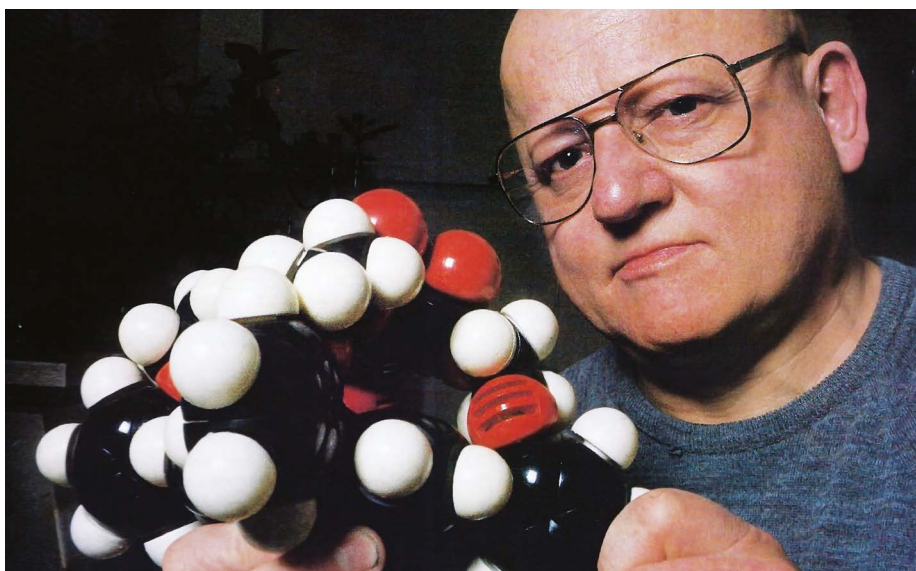
"Chemistry," Jack Baldwin once said in his direct way, "is about making forms of matter that have never existed." Baldwin was best known for formulating a set of rules that predict how likely it is that atoms (mostly carbon) in a synthesis will link into rings, a structural feature of many biological molecules and drugs. Published in just three pages (with a one-sentence abstract) in 1976 (*J. E. Baldwin J. Chem. Soc. Chem. Commun.* 734–736; 1976), Baldwin's rules have been fundamental to organic synthesis in the pharmaceutical and agrochemical industries, and to understanding biology from a chemical perspective. He died on 4 January, aged 81.

His passions also encompassed finding out how nature makes chemicals that researchers cannot. This led him to 'biomimetic' synthesis: using the principles of nature to improve the generation of biomolecules in the laboratory. He particularly relished the challenge of 'molecules from Mars', his term for natural products whose biosynthesis was baffling.

Baldwin's interest in rings led him to study antibiotics that contain a β -lactam ring, the best known of which is penicillin. He worked initially with Edward Abraham, who had been part of the team that developed penicillin and who went on to reveal the activity of broad-spectrum antibiotics known as cephalosporins. Baldwin uncovered the mechanistic basis of the enzyme action that catalyses the formation of the two rings at the heart of the penicillin molecule. Others have since found that related enzymes are involved in many biological processes, including how the human body responds to low levels of oxygen.

Baldwin was born in London, and studied chemistry at Imperial College London, where he also did a PhD. He was supervised by Derek Barton, a pioneer of conformational analysis – the idea that the reactivity of a molecule could predict its preferred 3D shape – who later won a Nobel prize. Barton had a major impact on Baldwin's career.

After four years on the staff at Imperial College, Baldwin spent more than a decade in the United States, working first at Pennsylvania State University in State College and then at the Massachusetts Institute of Technology (MIT) in Cambridge. With an able young team and the latest instruments, his MIT period was particularly productive. To develop a detailed picture of how atoms arrange themselves



ANDY LANE

during organic reactions, he combined theoretical and geometric considerations with structural information. His team obtained this using techniques such as nuclear magnetic resonance and X-ray crystallography. At MIT, he created a class of biomimetic molecule that reversibly binds oxygen when complexed with iron, just as haemoglobin does in the blood, and formulated his rules for ring formation. It was also where he met his future wife, Christine Franchi, who built a career in academic publishing.

In 1978, Baldwin was recruited to head the Dyson Perrins Laboratory at the University of Oxford, UK. As only the fourth person to hold the chair in organic chemistry since the laboratory opened in 1916, he transformed his discipline at Oxford, in terms of both scientific ambition and equipment. Baldwin brought with him researchers from his internationally diverse lab at MIT, and continued to recruit people with a wide range of backgrounds.

Many of his students, who knew him as 'J.E.B.', went on to lead research all over the world. The output of his lab was prodigious: he is an author on at least 700 papers. In 1988, he became the founding director of the Oxford Centre for Molecular Sciences, which he headed for 10 years. The centre helped to link physical and biological sciences in Oxford.

The pioneering role of Oxford scientists in the extraction, testing and structural analysis of penicillin during the 1940s inspired Baldwin's extensive work on trying to make the drug from scratch. His respect for the optimally

efficient process by which microbes produce the molecule – even now, most penicillin antibiotics continue to be produced through fermentation – led him further into the field of biomimetic synthesis. His favoured approach for building complex multi-ring structures was to try to mimic nature's strategy of making a relatively simple linear framework that is predisposed to react to give multiple rings in a single step. The 'molecules from Mars' he made using this approach included unusual alkaloids derived from marine sponges and rare rainforest plants.

Baldwin had little time for the academic conventions of Oxford: he spoke his mind, and could seem pugnacious in scientific debate. But his forceful leadership style belied a generosity in his treatment of junior colleagues. Wholly committed to research, he never sought seats on prestigious committees, although his distinction brought many honours, including a knighthood in 1997. He developed links with the chemical industry and championed its role in society, encouraging his students to pursue industrial careers. Aside from science, he enjoyed good food, fine wine, powerful motorbikes, fast cars and his dogs. After he retired in 2005, he continued to co-author publications until just months before his death.

Georgina Ferry is a science writer based in Oxford, UK. Her books include biographies of Dorothy Crowfoot Hodgkin, Max Perutz and John Sulston.

Comment



IVAN KASHINSKY/PANOS

A tropical forest in the Bribri Indigenous region of Costa Rica.

Adopt a carbon tax to protect tropical forests

Edward B. Barbier, Ricardo Lozano, Carlos Manuel Rodríguez & Sebastian Troëng

A levy on fossil fuels can support and restore ecosystems that help to stem climate change.

Deforestation must be stopped in tropical countries to tackle the existential threats of climate change and biodiversity loss. The vast majority of Earth's species are in the tropics; forests there have taken in much of the carbon added to the atmosphere by human activities. Safeguarding these forests is central to slashing greenhouse-gas emissions and meeting the internationally agreed United Nations Sustainable Development Goals (SDGs)¹.

Sadly, in tropical countries and internationally, investments are woefully inadequate in conservation, restoration and improving land management to protect biodiversity and ecosystem services – collectively called 'natural climate solutions'^{1,2}.

To plug this gap, we urge more countries that have tropical forests to adopt a tropical carbon tax – in South and Central America, Africa, Asia and the Pacific. This is a levy on fossil fuels that is invested in natural climate

Comment

solutions. Such a policy can reduce the use of oil, gas and coal and mobilize domestic funds for adaptation and mitigation.

Costa Rica and Colombia have done this. Our own analysis shows that, if 12 other countries roll out a tropical carbon tax similar to Colombia's, they could raise US\$1.8 billion each year between them to invest in natural habitats that benefit the climate (see Supplementary Information).

We call on governments, development banks, financial investors and non-governmental organizations to support those countries that need financial and technical help to implement this policy, and to ensure that the money raised is spent efficiently and effectively.

Twin threats

Almost one-quarter of the emissions caused by humans come from agriculture, forestry, fibre and livestock production³. It has been estimated that tropical deforestation can contribute as much to emissions as do some large nations (see go.nature.com/37gmwvy). If present trends continue, by 2050 the world will have lost a further area of tropical forest almost the size of India – 289 million hectares⁴. This could squander half of the remaining global carbon budget for limiting warming to 1.5 °C above pre-industrial levels⁴.

Meanwhile, more than three-quarters of species live in the tropics. These are under greater threat of extinction than is life elsewhere, mainly because of deforestation⁵.

There is a quick, cheap way to halt these trends: reducing the conversion of land in the tropics, especially of forests, peatlands and mangroves. Alongside cuts to fossil-fuel emissions, up to 37% of the mitigation needed to hold warming to the Paris agreement goal (to avoid the catastrophic impacts of climate change) might be achieved in this way, at a cost of less than \$100 per tonne of CO₂ equivalent¹ – the standard measure for greenhouse-gas emissions. One-third of these mitigation options could cost less than \$10 per tonne¹.

But ecosystem conservation, restoration and management received just 3% of global finance for climate mitigation in 2017–18: an average of \$18 billion⁶. Most of the remainder was spent on renewable-energy generation and on investments in low-carbon transport, such as railways and electric vehicles⁶.

Extra cash is unlikely to come from the international community in the near future, and aid and other funding is already scarce for biodiversity conservation in tropical countries². Such nations urgently need a new way to fund natural solutions to climate change.

Case studies

Colombia and Costa Rica have blazed a trail. Since 1997, Costa Rica has collected a 3.5% tax on fossil fuels. That now generates \$26.5 million per year⁷ (see go.nature.com/3jdpmk;

in Spanish). The tax was negotiated in Costa Rica's legislative assembly and supported by research from the non-governmental Tropical Science Center in San José, which examined the benefits of forests to the country's economy. Implementation faced little opposition because the tax was incorporated with other fiscal reforms. Surveys of fossil-fuel users indicated that they did not object if revenues were directed to forest conservation.

To invest the money raised, Costa Rica created its National Forest Fund (FONAFIFO). For example, from 1997 to 2018, the fund paid

“Investments in protecting biodiversity to reduce carbon emissions can favour poor people.”

out to landowners across 23.5% of the country – an area of 1.2 million hectares. They spent the money on projects to protect 1 million hectares of mature forest and 71,000 hectares under reforestation. The fund supports conservation of mature forests, reforestation using native or exotic species, and agroforestry systems that use a mix of trees and crops or grasslands. It has disbursed \$500 million to roughly 18,000 people, including those living across 162,000 hectares of Indigenous lands, such as the Cabécar and Bribri territories. Transparency and accountability of the fund's operations are important to its success and continued popularity, so strategic and operational plans, budgets, financial statements and other details are available online (see www.fonafifo.go.cr).

In the 1980s, Costa Rica had the highest deforestation rates in the world. Forest cover more than doubled between 1986 and 2013, rising to 53% (ref. 8). Although estimates remain uncertain, we think that the fossil-fuel tax, along with a decline in the profitability of livestock and the expansion of protected areas and ecotourism, contributed to this. The programme funded by the fuel tax has been especially effective away from protected areas and their buffer zones⁹.

Colombia rolled out a carbon tax in 2016 as part of sweeping fiscal reforms. These garnered broad political support because of the need to raise money for the country's peace process. The carbon tax was developed by the Ministry of Finance and Ministry of Environment and Sustainable Development, and is collected from companies producing or importing fossil fuels.

Colombia's tax of \$5 per tonne of emitted carbon yielded revenues of \$148 million in 2017 and \$91 million in 2018 (see go.nature.com/3b8ufkj; in Spanish). These go to the Colombian Peace Fund (*Fondo Colombia en Paz*), from which 25% is used to manage coastal erosion, reduce and monitor deforestation, conserve water



sources, protect strategic ecosystems and combat climate change. A further 5% is used to strengthen Colombia's National System of Protected Areas. The revenue will be used for conservation projects in the following prioritized areas: flood-plain forests, tropical montane cloud forests, tropical humid forests, tropical savannahs and Andean forests. These projects are in the development phase and are waiting to access the fund. There is also a project to enhance the Colombian Environmental Information System (SIAC), a web-based platform that provides official information on the state of the country's natural resources and which is under development (see go.nature.com/2hthzqw; in Spanish).

A mechanism called carbon neutrality allows companies to reduce their tax burdens by buying certified carbon credits from conservation and restoration projects in Colombia that adhere to internationally recognized standards. For example, a company might buy a credit in a region that promotes social initiatives with communities that are involved in managing these projects. This is the case for communities in the Chocó departmental



People in the Democratic Republic of the Congo at a charcoal market — the fuel is one of the causes of deforestation in the country.

FEDERICO SCOPPA/AFP/GETTY

region of northwestern Colombia, such as those living near towns including Acandí, El Carmen del Darién and Baudó.

Join in

Up to 70% of the world's biodiversity is found in just 17 'megadiverse' countries¹⁰. Thirteen contain tropical forests. In 2018, these countries lost almost 7.3 million hectares of forests — an area roughly the size of Panama. According to our estimates, that represented nearly 30% of global deforestation and may have released about 7% of worldwide carbon emissions (see Supplementary Information and www.globalforestwatch.org/map).

Two scenarios illustrate how these countries could benefit from a tropical carbon tax¹¹ (see also Supplementary Information). The first assumes that each follows a similar policy to that of Colombia, introducing a tax of \$5 per tonne of carbon emitted, and allocating 30% of the revenues to natural solutions to climate change and measures that conserve forests. The second assumes a tax of \$15 per tonne of carbon emitted and 70% allocation.

We provide this second option because

we think that both the urgency and interest in addressing climate change and biodiversity loss will continue to grow. It is also likely that some governments will choose to adopt such a higher carbon price and allocate more revenues to natural climate solutions.

For some countries, notably India, the Philippines, Mexico, Ecuador and Malaysia, the sums raised could provide hundreds of dollars per hectare to counter forest loss. The more ambitious policy could yield nearly \$13 billion each year for natural climate solutions.

Brazil, the Democratic Republic of the Congo and Indonesia would benefit the most, because they currently have the greatest amount of deforestation. Countries that have experience in developing high-quality carbon-offset projects, such as Peru and Ecuador, are well positioned to adopt a tropical carbon tax (see go.nature.com/2tptk21).

Politically challenging

There are three main criticisms of funding natural climate solutions through carbon taxes. First, that they cause 'leakage' — the displacement of deforestation to other areas.

Second, that they reduce the incentive to reduce emissions through renewable energy. And third, that the tax revenue should be used for other purposes.

We think that each of these problems can be addressed. National tax schemes reduce the likelihood of leakage in each country. Renewable-energy production and natural climate solutions are both essential, as indicated by scenarios from the Intergovernmental Panel on Climate Change³. Finally, although there are many worthy uses of tax revenue, the severity of climate change and biodiversity loss means that stemming both at once is a development priority for tropical-forest countries.

We also recognize that it can be politically challenging to introduce measures that increase the cost of living. But as the examples in Costa Rica and Colombia illustrate, investments in protecting biodiversity to reduce carbon emissions can favour poor people because such investments have wider social benefits beyond landowners and parks⁹. In Costa Rica, forests and high levels of poverty can often be found in the same districts, so



A cotton-top tamarin (*Saguinus oedipus*) in one of Colombia's protected national parks.

revenues destined for conservation can also contribute to social development. The Costa Rican government prioritizes such districts for payouts for ecosystem services. It also assists smallholder farmers and Indigenous communities in submitting requests for funds. Today, 40% of beneficiaries in Costa Rica are communities that live below the poverty line.

Ecosystem services such as drinking-water supply, food provision and cultural services are estimated to contribute between 50% and 90% of income and subsistence among the rural poor and those who live in forests¹². Such services can make an important contribution to ending extreme poverty (SDG 1), achieving zero hunger (SDG 2), improving health (SDG 3) and meeting many of the other 14 SDGs¹².

International support

The World Bank, the International Monetary Fund (IMF) and other multilateral agencies should encourage more countries to adopt a tropical carbon tax. The IMF already promotes carbon taxes as an efficient and fiscally responsible way of reducing emissions, with revenues being used for much-needed public investments in developing countries¹³. The international community can support more-widespread adoption of a tropical carbon tax in two important ways.

First, some tropical-forest countries and other low-income nations will require extra financial assistance because they might be unable to raise sufficient funds from a carbon tax. For example, if Papua New Guinea, Madagascar and the Democratic Republic of the Congo adopted Colombia's

approach for combating each hectare of forest loss with natural climate solutions, they would generate only \$23, \$3 and \$1 per hectare, respectively (see Supplementary Information). Top-up financing could come from bilateral assistance, or from the Special Climate Change Fund and the Least Developed Countries Fund. Both of these

“The severity of climate change and biodiversity loss means that stemming both at once is a priority.”

are managed by the Global Environmental Facility for the UN Framework Convention on Climate Change (UNFCCC).

Second, many tropical-forest countries will require technical support to guide and monitor their investments. Countries should comply with recognized global quality marks such as the Verified Carbon Standard (<https://verra.org/project/vcs-program>) and the Climate, Community and Biodiversity Standard (<https://verra.org/project/ccb-program>). The first is the world's most widely used voluntary programme for mitigating greenhouse-gas emissions. The second identifies projects that simultaneously address climate change, support local communities and smallholders, and conserve biodiversity. Currently, the projects that have been validated and verified encompass more than 10 million hectares, an area the size of Iceland (see <https://verra.org/project/ccb-program>).

Tropical countries are already showing interest in carbon-pricing initiatives and

natural climate solutions. Next week, Costa Rica will host a high-level meeting on the subject in San José with government and business leaders from Peru, Ecuador, Mexico and Chile, as well as Colombia.

And several major international events in 2020 provide a platform for supporting global action towards a tropical carbon tax. These include the International Union for Conservation of Nature's World Conservation Congress in June, the 15th meeting of the Conference of the Parties (COP15) to the Convention on Biological Diversity in Kunming, China, in October, and the 26th session of the UNFCCC Conference of the Parties (COP26) in Glasgow, UK, in November. We suggest that, at these meetings, policymakers explicitly highlight and incorporate a tropical carbon tax in agreements and decisions.

Tropical deforestation and land-use change must be halted to safeguard the climate and global biodiversity. The widespread adoption of a tropical carbon tax is a practical way forward.

THOMAS MAREN/NPL

The authors

Edward B. Barbier is a university distinguished professor in the Department of Economics, Colorado State University, Fort Collins, Colorado, USA. **Ricardo Lozano** is Minister of Environment and Sustainable Development, Colombia. **Carlos Manuel Rodríguez** is Minister of Environment and Energy, Costa Rica. **Sebastian Troeng** is executive vice-president of Conservation International in Arlington, Virginia, USA. e-mail: edward.barbier@colostate.edu

1. Griscom, B. W. et al. *Proc. Natl Acad. Sci. USA* **114**, 11645–11650 (2017).
2. Barbier, E. B., Burgess, J. C. & Dean, T. J. *Science* **360**, 486–488 (2018).
3. Intergovernmental Panel on Climate Change. *Special Report on Climate Change and Land* (eds Shukla, P. R. et al.) (IPCC, 2019).
4. Busch, J. & Engelmann, J. *Environ. Res. Lett.* **13**, 015001 (2017).
5. Barlow, J. et al. *Nature* **559**, 517–526 (2018).
6. Climate Policy Initiative. *Global Landscape of Climate Finance 2019* (CPI, 2019).
7. National Fund for Forest Finance (FONAFIFO). *2019 Budget Plan* [In Spanish] (FONAFIFO, 2018).
8. Costa Rican Ministry of Environment and Energy. *Report on the State of the Environment 2017* [In Spanish] (2018); available at <https://go.nature.com/3b9ynev>
9. Robalino, J., Sandoval, C., Barton, D. N., Chacon, A. & Pfaff, A. *PLoS ONE* **10**, e0124910 (2015).
10. Mittermeier, R. A., Robles-Gil, P. & Mittermeier, C. G. (eds) *Megadiversity. Earth's Biologically Wealthiest Nations* (CEMEX/Agrupación Sierra Madre, 1997).
11. Boden, T. A., Marland, G. & Andres, R. J. *Global, Regional, and National Fossil-Fuel CO₂ Emissions (1751–2014)* (v. 2017) https://doi.org/10.3334/CDIAC/00001_V2017 (US Department of Energy, 2017).
12. Convention on Biological Diversity. *Biodiversity and the 2030 Agenda for Sustainable Development: Technical Note* (CBD, 2019).
13. International Monetary Fund. *Fiscal Monitor: How to Mitigate Climate Change* (IMF, 2019).

Supplementary Information accompanies this article: see go.nature.com/3twakvj.

Correspondence

Clamp down on trade-ban violations

In response to the deadly outbreak of coronavirus 2019-nCoV (see *Nature* <http://doi.org/dk47>; 2019), China has temporarily banned the sale of wildlife in markets, restaurants and online. Given that much of this trade is already illegal, stricter enforcement and prosecution measures are needed if the consumption of wild animals is to be brought under control.

At present, prosecutions are often obstructed because of inconsistencies in the naming of species (Z.-M. Zhou *et al.* *Nature* 525, 187; 2015). Online trading in low-profile illegal wildlife as pets is commonplace (Y.-C. Ye *et al.* *Conserv. Sci. Pract.* <http://doi.org/dk49>; 2020). And the public's desire for exotic wildlife products remains undiminished – particularly for use in traditional medicines. Dodging the law on such a scale is a disaster for global biodiversity and animal welfare, as well as for human health.

When, or if, wildlife trade is again permitted, it must be better scrutinized so that stringent hygiene and quarantine standards at markets can be enforced. Advertisements will need to include the scientific names of species as well as their provenance. Supplies from licensed captive breeders must be properly regulated and inspected – a step that would also help pin down violations of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES).

Zhao-Min Zhou* China West Normal University, Nanchong, China.
zhouzm81@gmail.com
*On behalf of 4 correspondents (see go.nature.com/39iz5hc).



The Chinese authorities have imposed a temporary ban on the wild-food trade.

Total ban on wildlife trade could fail

The Chinese government's temporary ban on the domestic transport and sale of wild animals following the emergence of coronavirus 2019-nCoV is welcomed by environmental non-governmental organizations pushing for a permanent ban (see go.nature.com/3b9kqcx). But China's cultural demand for wildlife items could mean that a blanket ban would be counterproductive.

Total bans are controversial because they risk fuelling an intractable, uncontrolled and highly priced illegal trade, sustained by the rising incomes and social status of the country's growing middle class (D. W. S. Challender *et al.* *Front. Ecol. Environ.* 17, 199–200; 2019). China's complex culture is at the root of its demand for exotic wildlife items such as pangolin scales, tiger bones and rhino horns. Likewise, the consumption of game meat is regarded as healthy as well as an indicator of wealth. Markets selling such produce are prime candidates for passing

on new viruses.

This complex issue needs to be managed through initiatives that discourage consumption, such as wisely directed education campaigns that aim to discredit engrained cultural beliefs.

Joana Ribeiro* CIBIO-InBIO, University of Porto, Portugal.
joanateixeiraribeiro@gmail.com
*On behalf of 4 correspondents (see go.nature.com/2udauk9)

Romania: help astronomers return

The political climate seems to be improving under the new government in Romania, but the country's research is still hampered by the Romanian Academy's outdated regulations. These discourage Romanian citizens who have pursued careers abroad from returning to many institutes – including to the Astronomical Institute of the Romanian Academy (AIRA) in Bucharest. As an astronomer of Romanian origin working in Spain, I urge the government to persuade the Romanian Academy to reform its regulations and open up its

research to its citizens working abroad and to scientists from the rest of the European Union.

There are no graduate astronomy departments in Romanian universities and the country has no useful observatories. When senior astronomers retire, there is no one to replace them because the bright young astrophysicists have all decamped abroad.

Although the academy announced in 2016 that its doors are open to EU researchers, jobs are advertised only in Romanian. Researchers wishing to return home must have their foreign PhD qualifications validated in Romania; they are then graded according to their previous Romanian employment.

Foreign candidates and citizens who trained abroad are excluded from senior research positions. For example, a high-grade post in astrophysics recently went to a home-grown researcher from another discipline. The academy's arcane rulings must be scrapped if Romania is to compete in international science.

Ovidiu Vaduvescu La Palma, Canary Islands, Spain.
ovidiu.vaduvescu@gmail.com

News & views

Applied physics

Quantum cascade laser lives on the edge

Sunil Mittal & Edo Waks

Devices known as quantum cascade lasers produce useful terahertz radiation, but are typically highly sensitive to fabrication defects. This limitation has now been overcome using a property called topological robustness. **See p.246**

Electromagnetic waves with frequencies in the terahertz range (300 GHz to 10 THz) have applications in many areas, from imaging and security screening to the atmospheric and biological sciences. Semiconductor devices called quantum cascade lasers (QCLs) provide the most compact and efficient way to generate terahertz radiation. In QCLs, electrons cascade down in energy through a series of discrete quantum energy levels, emitting a photon at each step¹. But, as with all compact semiconducting lasers, QCLs are notoriously sensitive to fabrication imperfections, which results in device-to-device variability of the laser output frequency. Now, on page 246, Zeng *et al.*² report the realization of a terahertz QCL that is insensitive to such disorder. This achievement opens the door for terahertz lasers and optoelectronics that have unprecedented stability and fabrication reproducibility.

Lasers use a process known as optical feedback to build up light intensity and stimulate electrons to emit photons. A common way to introduce this feedback uses a structure called an optical cavity, which is typically composed of mirrors that reflect the emitted light back into the device. Compact lasers, however, use more-complex structures such as photonic crystals – materials that have a periodically varying refractive index. If this periodicity is carefully engineered, photonic crystals can be used to reflect light waves of only the desired frequency, and so achieve lasing³. But this approach is highly sensitive to disorder, because any imperfections in the photonic crystal cause reflections that result in waves of unwanted frequencies. These compete with the desired waves, leading to unstable light intensity and poor laser efficiency.

In the past few years, ‘topological’ photonic structures have emerged as a way to make

photonic devices that are insensitive to disorder. This area of research originated from concepts developed in condensed-matter physics. Over the past two decades, condensed-matter physicists have been able to use the mathematical descriptions of symmetries and topology to characterize different forms of matter. Of particular relevance to the current work are exotic materials known as topological insulators⁴.

As the name suggests, these materials are insulators – that is, they do not conduct electricity in their interior. However, they host conducting electronic states at their boundaries. Such edge states can carry current in only one direction and are therefore robust against disorder that would otherwise scatter charge carriers. This robustness of

edge states is a manifestation of the overall topological properties of the material. Topological insulators are so insensitive to disorder that they were previously used to define the unit of resistance: the ohm.

Although topological physics originated in the field of electronics, it has begun to inspire photonics⁵. Disorder and scattering are even more problematic in optics than in electronics, because photons exhibit strong interference effects that can lead to complicated, difficult-to-control laser behaviour. Translating topological protection into the optical domain opens up the possibility of making robust optical systems. In particular, topological lasers can emit light in a way that is robust against scattering and other consequences of imperfections. But previous realizations of topological lasers^{6–8} have operated at frequencies above the terahertz range.

Zeng and colleagues overcame this limitation by incorporating topological protection into a QCL. To achieve this, they used a topological model known as the valley Hall effect, which relies on breaking the spatial-inversion symmetry of a crystal lattice⁹ (its symmetry under the combination of a 180° rotation and a mirror reflection). Specifically, the authors used a gallium arsenide–aluminium gallium arsenide substrate as the gain material – the medium in which light is amplified. This substrate contained layered semiconductor structures called quantum wells that were designed to support quantum cascade lasing.

The authors drilled a triangular lattice of holes in the gain material (Fig. 1). The symmetries of

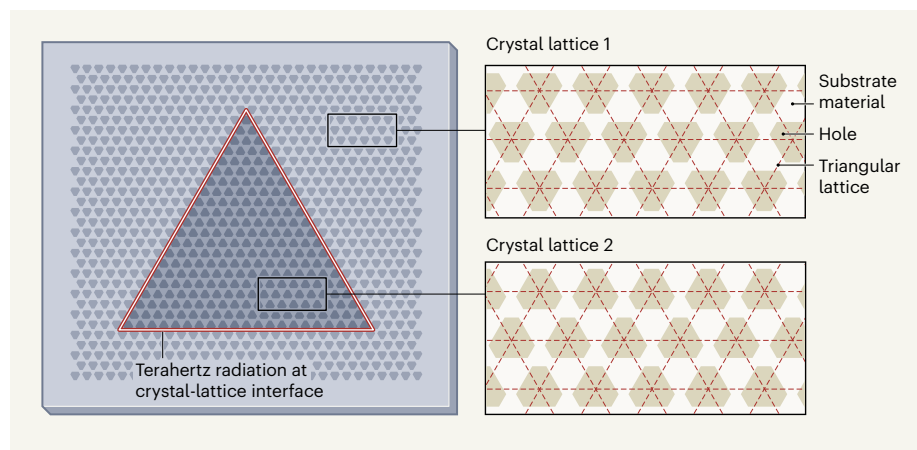


Figure 1 | Design of a topological laser. Zeng *et al.*² have made a laser in which terahertz radiation is emitted from the interface between two triangular crystal lattices that consist of quasi-hexagonal holes in a substrate material. The crystal lattices are topologically inequivalent because the orientation of the holes is flipped in one lattice with respect to the other, and this leads to the emergence of exotic photonic states called edge states at the crystal-lattice interface. The topological nature of these edge states renders the laser robust against fabrication imperfections.

this lattice resulted in the emergence of two valleys in the energy–momentum band structure – the relationship between the energy and momentum of photons in the material. The authors made the holes quasi-hexagonal so that they broke the spatial-inversion symmetry of the lattice and rendered the two valleys topologically inequivalent. This led to the formation of topological edge states at the interface between two such crystal lattices in which the orientation of holes (and valleys) was flipped in one lattice with respect to the other.

Zeng and co-workers used these topological edge states to design and make a robust ring resonator (a type of optical cavity that traps light at certain ‘resonance’ frequencies) in the form of a triangle (Fig. 1). It is this triangular cavity that, along with the light amplification from the substrate material, forms a topological laser. The laser produces light of many frequencies that are separated by similar frequency gaps. These frequencies correspond to the resonance frequencies of the triangular cavity and fall within the frequency range of the QCL gain material.

The authors measured light emission from different points along the perimeter of the cavity and discovered that the emission at each point had the same resonance frequencies. This indicates that these waves travelled through the length of the cavity, traversing the sharp (60°) bends at the corners of the triangle. Furthermore, Zeng *et al.* found that the lasing frequencies did not change when they introduced defects, in the form of extra holes, around the cavity, demonstrating the robustness of the QCL.

Another key feature of this laser is that energy is ‘pumped’ into the device electrically. Previous topological lasers^{6–8} have been optically pumped, which means that they require a second laser source to drive the topological laser to generate light. This pumping scheme severely limits practical applications. However, similar to many commonly used lasers (such as laser pointers), Zeng and colleagues’ QCL can be directly driven by an electrical current, allowing it to be powered, in principle, by a battery or a wall outlet, rather than by another laser.

Robustness against defects and disorder is one defining characteristic of topological physics, but another important feature is a type of asymmetry called chirality. In particular, in the valley Hall effect, the two valleys are associated with photons of opposite circular polarization in the plane of the material. If right-circularly polarized photons travel to the left, then left-circularly polarized ones would travel to the right. Realizing this chirality represents a crucial future step towards terahertz topological lasers in which light waves flow around a ring resonator in only one direction. The chirality could be incorporated either by

explicitly breaking time-reversal symmetry (a symmetry in which reversing the direction of light waves is equivalent to running time backwards) or by introducing directional light amplification in the cavity.

Zeng and co-workers’ results pave the way for studying topology in a previously inaccessible part of the electromagnetic spectrum. One area of great interest for future research is the application of other topological models, such as exotic (higher-order) topological insulators, to make robust terahertz lasers that have other geometries. For example, these lasers could emit light at the corners, rather than at the edges, of a triangular cavity.

Another fascinating prospect is the exploration of non-Hermitian (open) physical systems at terahertz frequencies, in which the presence of light amplification and loss can lead to the emergence of features such as parity–time symmetry (symmetry under the combination of a mirror reflection and time reversal) and exceptional points (spectral features

that correspond to coalescing resonances)¹⁰. The realization of topological photonics in the terahertz range could therefore serve as a catalyst for the development of practical devices, and also enable a better fundamental understanding of topological physics and complex (nonlinear) optoelectronics.

Sunil Mittal and **Edo Waks** are at the Joint Quantum Institute, University of Maryland, College Park, Maryland 20742, USA.
e-mails: mittals@umd.edu;
edowaks@umd.edu

1. Faist, J. *et al.* *Science* **264**, 553–556 (1994).
2. Zeng, Y. *et al.* *Nature* **578**, 246–250 (2020).
3. Baba, T. *Nature Photon.* **2**, 465–473 (2008).
4. Hasan, M. Z. & Kane, C. L. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
5. Ozawa, T. *et al.* *Rev. Mod. Phys.* **91**, 015006 (2019).
6. Bandres, M. A. *et al.* *Science* **359**, eaar4005 (2018).
7. Bahari, B. *et al.* *Science* **358**, 636–640 (2017).
8. St-Jean, P. *et al.* *Nature Photon.* **11**, 651–656 (2017).
9. Ma, T. & Shvets, G. *New J. Phys.* **18**, 025012 (2016).
10. El-Ganainy, R. *et al.* *Nature Phys.* **14**, 11–19 (2018).

Structural biology

Long-distance coupling in a promiscuous protein

Thorsten Althoff & Jeff Abramson

Unlike many sugar-transporting proteins, a transporter in one species of malaria parasite can import several types of sugar equally effectively, aiding the parasite’s survival. The structure of this protein reveals the reason for its versatility. **See p.321**

Most cases of malaria are caused by the protozoan parasite *Plasmodium falciparum*¹. Given that there are more than 400,000 malaria-associated deaths annually, and that *P. falciparum* is constantly evolving to resist pharmacological therapies, opportunities for developing drugs that target this organism must be continuously explored. A protein called the *P. falciparum* hexose transporter 1 (PfHT1) has a proclivity for scavenging sugars from an infected host’s red blood cells to improve the parasite’s chances of survival in these cells, and is therefore a drug target. On page 321, Qureshi *et al.*² describe the 3D structure of PfHT1, and identify a mechanism that couples the docking of a sugar in the PfHT1 binding site to the process by which sugars are gated through the protein. This coupling facilitates the protein’s substrate promiscuity – that is, its ability to transport a wide range of sugar molecules effectively, a feature that gives the parasite a distinct survival advantage.

Transporter proteins shuttle substrate

molecules across the otherwise impermeable lipid bilayer of the cell membrane. The functional and dynamic properties of these membrane-embedded proteins are fundamentally related to their 3D structures, which are modulated at the atomic level over a broad range of timescales. Membrane transporters use the alternating-access mechanism for gating³, in which access to the substrate-binding site switches from one side of the membrane to the other (Fig. 1).

The development of methods for determining the structures of membrane proteins in the past few years has produced near-complete pictures of the translocation mechanisms of several classes of transporter – that is, the global rearrangements that the proteins undergo during translocation cycles of substrate binding, transport and release have been visualized at atomic resolution. Intuitively, the substrate specificity of transporters has generally been found to depend on the amino-acid residues at the binding

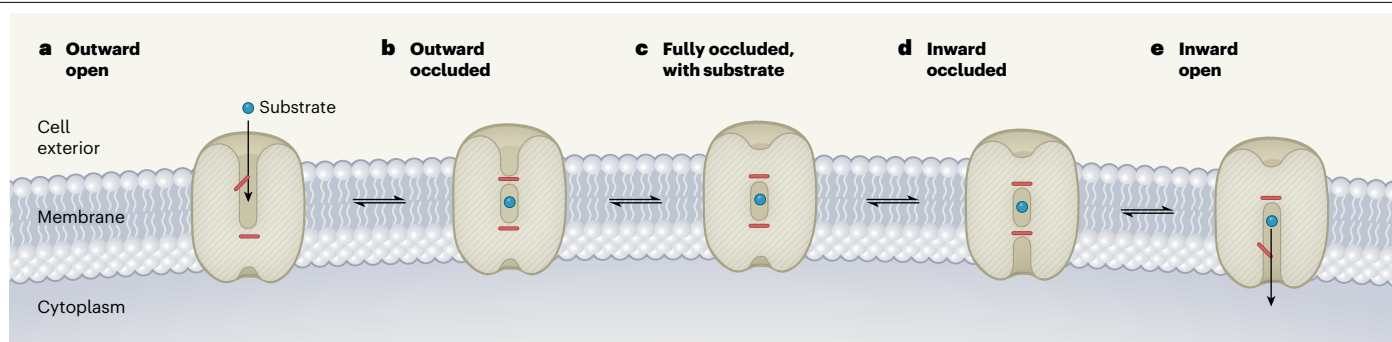


Figure 1 | The alternating-access mechanism. Transporter proteins facilitate the passage of substrate molecules across cell membranes. Access to the substrate-binding site in the middle of transporters is controlled by two gates (red). **a**, In the outward open state, a pathway from the cell exterior allows substrates into the protein. **b**, In the outward occluded state, a substrate is trapped between the gates, but the outward-facing pathway is still present. **c**, In the fully occluded state with a bound substrate, no pathways are available. **d**, In the inward occluded

state, a pathway to the cytoplasm has formed, but the gate remains closed. **e**, In the inward open state, substrates can exit to the cytoplasm. Qureshi *et al.*² report the structure of PfHT1, a sugar transporter from the malaria parasite *Plasmodium falciparum*. They find that the binding of a sugar substrate to the structure shown in **a** is coupled to the gating mechanism, and that the transition from **a** to **c** occurs much faster than in other sugar transporters. This explains why PfHT1 transports a wide range of sugar molecules equally effectively, unlike other sugar transporters.

site. The structure of PfHT1 now implies that another mechanism affecting substrate specificity might be at play.

Red blood cells infected by *P. falciparum* consume about 100 times more glucose than do non-infected cells⁴ because the parasite continuously metabolizes sugars from these cells to support its growth and replication. Because PfHT1 is responsible for transporting sugars from host cells, it has a crucial role in supporting this metabolism. It belongs to the well-studied major facilitator superfamily (MFS) of transporters, which promote the diffusion of substrates across the cellular membrane. It has the same overall 3D structure as the distantly related human GLUT transporters⁵. But whereas these specialize in the transport of either D-glucose or D-fructose, PfHT1 transports both of these sugars, and some others, with comparable efficiency.

Qureshi *et al.* resolved the 3D structure of PfHT1 in which D-glucose is captured in the sugar-binding site, and found that the protein was in a fully occluded conformation – that is, the transporter protein completely shielded the sugar from the aqueous environments on either side of the cell membrane. The structure therefore provides a snapshot of the substrate during a part of the translocation cycle that had not previously been visualized for an MFS transporter.

Armed with their structure, the authors carried out extensive transport studies to try to work out why PfHT1 has less substrate selectivity than its human GLUT counterparts. They first demonstrated that the same set of amino-acid residues in PfHT1 is required to bind D-glucose and D-fructose. They then replaced residues in and around the sugar-binding site of PfHT1 by residues found in GLUT transporters, but none of these mutations conferred GLUT-like selectivity on the resulting proteins. They thus concluded that the unusual lack of selectivity of PfHT1 cannot be explained on the basis

of the sugar-binding residues alone.

So how can the substrate promiscuity of PfHT1 be explained? It has been known since the first structures of MFS transporters were reported^{6,7} in 2003 that bundles of α -helices in the proteins ‘rock’ around the central substrate-binding site, thereby establishing the alternating pathways for substrates through the protein: an outward-facing pathway, which allows substrates into the transporter from the cell exterior, and an inward-facing pathway that allows substrates to enter the cytoplasm (Fig. 1). By considering their structure of the fully occluded state of PfHT1 alongside structures of other sugar transporters captured at different stages in the translocation of D-glucose^{8–13}, Qureshi *et al.* were able to describe a complete translocation cycle.

The authors found that, surprisingly, all of the sugar-binding residues maintain their orientations throughout the cycle. This implies that the switches from the outward-facing conformation of PfHT1 to the fully occluded state, and then to the inward-facing conformation, are not driven by structural rearrangements at the sugar-binding site. Instead, they are driven by a previously unknown mechanism.

Qureshi and co-workers’ analysis of the gating mechanism of PfHT1 revealed interactions involving hydrophilic amino-acid residues in two transmembrane α -helices in the occluded state. By contrast, in human GLUT proteins, the equivalent residues are larger and more hydrophobic. Experiments in which the authors substituted these gating residues in PfHT1 with other residues demonstrated that they are crucial for sugar transport. Notably, the gating residues are about 15 Å away from the sugar-binding site – a large distance. This indicates that the binding of a sugar is coupled to remote conformational changes associated with gating of the transporter, a type of mechanism known as allosteric coupling. Thus, the ability of PfHT1, unlike its human counterparts, to transport

many similar substrates results from its substrate-driven gating dynamics, which allows it to adopt the occluded conformation more easily and rapidly.

The authors also carried out experiments to investigate how PfHT1 is inhibited by two small-molecule antimalarial drugs (C3361 and MMV009085). This allowed them to identify a hydrophobic pocket in the transporter that probably facilitates the binding of inhibitory drug molecules, and that might help to guide the design of new antimalarial compounds. However, the most exciting finding is the allosteric coupling between substrate binding and gating – it suggests that substrate recognition in transporters can be a consequence of the transporter’s conformational dynamics, rather than being the result of protein–substrate interactions, which underpin the conventional ‘lock and key’ model of how molecules interact with their biological targets.

Thorsten Althoff and Jeff Abramson are in the Department of Physiology, University of California, Los Angeles, Los Angeles, California 90095, USA. e-mail: jabramson@mednet.ucla.edu

1. World Health Organization. *World Malaria Report 2019* <https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019> (2019).
2. Qureshi, A. A. *et al.* *Nature* **578**, 321–325 (2020).
3. Jardetzky, O. *Nature* **211**, 969–970 (1966).
4. Roth, E. Jr *Blood Cells* **16**, 453–466 (1990).
5. Woodrow, C. J., Burchmore, J. R. & Krishna, S. *Proc. Natl Acad. Sci. USA* **97**, 9931–9936 (2000).
6. Abramson, J. *et al.* *Science* **301**, 610–615 (2003).
7. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D.-N. *Science* **301**, 616–662 (2003).
8. Deng, D. *et al.* *Nature* **526**, 391–396 (2015).
9. Nomura, N. *et al.* *Nature* **526**, 397–401 (2015).
10. Deng, D. *et al.* *Nature* **510**, 121–125 (2014).
11. Sun, L. *et al.* *Nature* **490**, 361–366 (2012).
12. Quistgaard, E. M., Löw, C., Moberg, P., Trésaugues, L. & Nordlund, P. *Nature Struct. Mol. Biol.* **20**, 766–768 (2013).
13. Wisedchaisri, G., Park, M.-S., Iadanza, M. G., Zheng, H. & Gonen, T. *Nature Commun.* **5**, 4521 (2014).

This article was published online on 29 January 2020.

Ordered absences out of the blue

Adam Jaffe & Jeffrey R. Long

Prussian blue analogues are archetypes of coordination solids, in which metal ions are bridged by ligands to form extended network structures. An analysis reveals a surprising ordering of the gaps found in their crystal lattices. **See p.256**

The centuries-old pigment Prussian blue and its analogues are archetypes of compounds known as coordination solids, and have had an unparalleled role in advancing our understanding of inorganic chemistry and materials^{1,2}. The wide-ranging structural, electronic, magnetic and optical properties of Prussian blue analogues (PBAs) have been repeatedly leveraged towards applications that include energy storage³, catalysis⁴, ion trapping⁵ and gas storage⁶. However, studying the surprisingly complex atomic-scale structures of PBAs remains a long-standing challenge. On page 256, Simonov *et al.*⁷ report that they have successfully grown single crystals of PBAs, which have previously been notoriously elusive. By coupling X-ray measurements of the crystal lattices with a simple but effective theoretical model, the authors reveal an unexpected ordering of vacancies – absent nodes in the lattices that correspond to missing metal–anion units. This structural insight could enable yet another means of adjusting the properties of these extraordinary materials.

Prussian blue ($\text{Fe}_4[\text{Fe}(\text{CN})_6]_3 \cdot 14\text{H}_2\text{O}$) was first reported⁸ in 1710 and was widely used as a deep-blue pigment. The eventual determination of its crystal structure greatly expanded the conceptual boundaries of inorganic chemistry. X-ray diffraction experiments performed on powders⁹, and later on single crystals¹⁰, of Prussian blue revealed the parent structure shared by all PBAs: a cubic framework in which two different types of metal cation act as ‘nodes’ linked in three dimensions by cyanide anion (CN^-) ‘struts’ (Fig. 1a). PBAs therefore have the general formula $\text{M}[\text{M}'(\text{CN})_6]_x$, in which M and M' are chemically distinct metal ions; the $[\text{M}'(\text{CN})_6]^{3-/4-}$ complex ion unit (Fig. 1b) is known as a hexacyanometallate ion, and carries either three or four negative charges. The study of the PBA parent structure enriched our fundamental understanding of the coordination chemistry of transition metals (how ligand molecules or ions bind to transition-metal ions such as iron, cobalt and copper), and

demonstrated that coordination solids that have multidimensional connectivity can act as porous framework materials through which molecules and ions can move.

The idealized crystal structures of PBAs correspond to the cubic framework described above, but belie a hidden degree of complexity that is crucial in determining their physical properties. The true atomic-scale structures contain vacancies corresponding to absent hexacyanometallate ions (Fig. 1b), which form pores that are typically filled with water molecules. The concentration and ordering (networking) of vacancies control the pathways through which mass can move within the materials, and can therefore tune the ability of PBAs to reversibly transport ions or small molecules. Insight into how vacancy ordering is affected by the chemistry of PBAs, or by the conditions used to synthesize them, can thus provide guidelines on how to

tailor the properties of these compounds for applications.

X-ray-scattering measurements on PBA powders, beginning with the early diffraction studies on Prussian blue⁹, yielded structural information for these compounds. But the random orientation of millions of crystallites in powders leads to loss of information that is retained if measurements are performed on single crystals. To gain this extra insight and illuminate vacancy behaviour, Simonov *et al.* sought to produce crystals of a series of PBAs that contained different metal-ion combinations. Growing single crystals of PBAs is challenging because of the rapidity with which microcrystalline powders precipitate when solutions of PBA precursors are combined. However, the authors found that controlled mixing of these solutions over the course of weeks produced single crystals suitable for X-ray-scattering analysis.

Simonov and co-workers observed clear indicators of non-random ordering of vacancies in the scattering data for their PBA crystals. This ordering depends on each crystal's chemical composition and the conditions used to crystallize it. To understand the diversity of the vacancy networks, the authors developed a simple two-part model to simulate vacancy ordering. The model considers only the trade-off between the preference of these compounds to adopt a uniform vacancy distribution, and the preference for lattice sites to have a certain local symmetry, yet it effectively reproduces the experimental X-ray scattering results.

Notably, the authors' insights enable the vacancy-network architectures of PBAs to be predicted by considering only a few factors

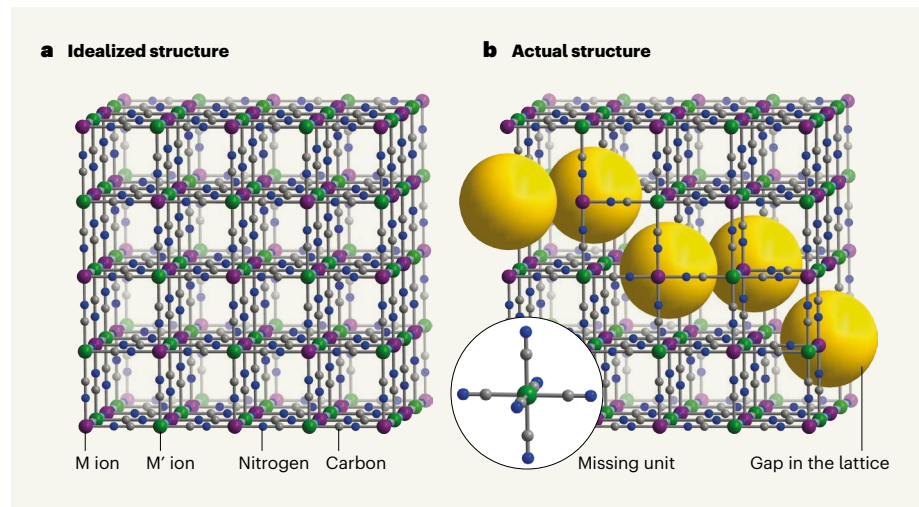


Figure 1 | Vacancies in Prussian blue analogues. **a**, Compounds known as Prussian blue analogues (PBAs) have the formula $\text{M}[\text{M}'(\text{CN})_6]_x$, where M and M' are two chemically distinct metal atoms. The idealized crystal structure of a PBA is a cubic framework in which M and M' ions act as ‘nodes’ connected by cyanide ions (CN^-), which act as ‘struts’. **b**, The actual crystal structures contain vacancies – gaps in the lattice that correspond to missing $[\text{M}'(\text{CN})_6]^{3-/4-}$ units. Networks of vacancies can form pathways that allow molecules or ions to be transported through PBAs, a potentially useful characteristic. Simonov *et al.*⁷ have used X-ray measurements of single crystals of PBAs and numerical modelling to reveal the hidden order of vacancies in PBAs.

that depend on the two model parameters, such as the choice of metal, precursor concentrations and the temperature of crystallization. Some networks turn out to have relatively direct pathways through which a molecule or ion could move, whereas other networks' pathways are more tortuous. By selecting PBAs that have direct pathways facilitating mass transport, these materials can be optimized for use as battery electrodes, catalysts or ion-exchange materials.

Simonov and colleagues' work addresses a long-standing lack of detailed knowledge about the structural vacancies that determine the physical properties of Prussian blue and its analogues. But numerous challenges remain before the predictive potential of their results can be fully realized. Although remarkably effective, the modelling analysis does not consider further possible complexities, such as the effects of ionic species that dwell in the PBA pores. Extrapolation of the findings from these single-crystal studies to powder samples, which are more technologically relevant, will require further challenging experiments and enhanced modelling that considers the surface structure and chemistry of micro-particles. Great care will also be needed to work out how each of the variables in a PBA synthesis correlate with the resulting vacancy ordering and material properties.

Although these challenges necessitate substantial further work, they also represent an opportunity to exert even greater control over the properties of PBAs, guided by a deeper understanding of structure–property relationships. Refinement of more-complex models will dictate how to take advantage of the many variables of a PBA synthesis. Not only has this work resulted in new-found control over the optimization of PBAs for applications in energy storage, ion capture and catalysis, but it also represents a platform on which to build a similar understanding of other framework materials, such as zeolites¹¹ and metal–organic frameworks¹², which have their own sets of challenges and promising applications.

Adam Jaffe and **Jeffrey R. Long** are in the Department of Chemistry, University of California, Berkeley, Berkeley, California 94720, USA. **J.R.L.** is also in the Department of Chemical and Biomolecular Engineering, University of California, Berkeley, and in the Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley. e-mails: adamjaffe@berkeley.edu; jrlong@berkeley.edu

1. Sharpe, A. G. in *The Chemistry of Cyano Complexes of the Transition Metals* (eds Maitlis, P. M., Stone, F. G. A. & West, R.) 1–302 (Academic, 1976).
2. Dunbar, K. R. & Heintz, R. A. in *Progress in Inorganic Chemistry* Vol. 45 (ed. Karlin, K. D.) 283–391 (Wiley, 1996).
3. Song, J. et al. *J. Am. Chem. Soc.* **137**, 2658–2664 (2015).
4. Kruper, W. J. Jr & Swart, D. J. US patent 4,500,704 (1985).
5. Kawamoto, T. et al. *Synthesiology Eng. Ed.* **9**, 139–154 (2016).

6. Kaye, S. S. & Long, J. R. *J. Am. Chem. Soc.* **127**, 6506–6507 (2005).
7. Simonov, A. et al. *Nature* **578**, 256–260 (2020).
8. Frisch, J. L. *Miscellanea Berolinensia ad incrementum scientiarum* **1**, 377–378 (1710).
9. Keggin, J. F. & Miles, F. D. *Nature* **137**, 577–578 (1936).

10. Buser, H. J., Schwarzenbach, D., Petter, W. & Ludi, A. *Inorg. Chem.* **16**, 2704–2710 (1977).
11. Baerlocher, C. et al. *Nature Mater.* **7**, 631–635 (2008).
12. Trickett, C. A. et al. *Angew. Chem. Int. Edn* **54**, 11162–11167 (2015).

Neurodegeneration

A protein's structure used to diagnose disease

Juan Atilio Gerez & Roland Riek

Parkinson's disease and multiple system atrophy involve the protein α -synuclein. Proof that aggregated α -synuclein adopts a different structure in each case suggests that its conformation underlies the distinct disorders. **See p.273**

A snowflake begins life as a tiny crystal that acts as a seed on which water molecules aggregate, increasing the size of the snowflake as it descends to earth. Proteins can also act as seeds – for instance, in a class of age-related disorders called amyloid diseases, in which thousands of copies of a type of protein known as an amyloid adopt an abnormal structure and aggregate in harmful clumps. In Parkinson's disease, aggregates of the amyloid protein α -synuclein accumulate in neurons. A rarer neurodegenerative disease, multiple system atrophy (MSA), involves α -synuclein aggregates in neuron-supporting cells called glia. It can be difficult to distinguish between the two disorders, given their overlapping symptoms, but they require different treatments. Shahnawaz *et al.*¹ provide an explanation for this difference on page 273: like two dissimilar snowflakes composed of identical water molecules, α -synuclein aggregates form distinct 3D architectures in each disease.

In vitro and animal experiments have previously indicated that different aggregate structures of α -synuclein, called strains, yield different effects². The various α -synuclein strains not only can have distinct cell-killing abilities and different seeding and propagation properties, but also can target different cell types and areas of the mammalian brain^{3,4}.

Shahnawaz *et al.* built on these previous findings using a technique called protein misfolding cyclic amplification (PMCA), which amplifies small amounts of α -synuclein aggregate, allowing thorough examination of minuscule samples. An amyloid-specific fluorescent dye is incorporated into the newly formed aggregates, enabling their analysis.

Impressively, the authors amplified and analysed samples from the cerebrospinal fluid of more than 200 people who had either Parkinson's disease or MSA, or who

were healthy (Fig. 1). They found that samples taken from people with Parkinson's disease displayed more fluorescence than those from people with MSA. Thus, PMCA could be used to discriminate between Parkinson's disease and MSA.

The different levels of fluorescence suggested that the amyloid dye interacted with each α -synuclein aggregate differently, and that distinct α -synuclein strains are involved in the two diseases. The authors confirmed this result by showing that the two strains could also be distinguished by using proteinase K digestion (an enzymatic treatment that breaks down strains that have different structures in different ways), and through other biophysical characterizations, including a microscopy approach called cryo-electron tomography.

Shahnawaz and colleagues' work has two major implications. First, it demonstrates that PMCA can be used as a diagnostic tool to discriminate between diseases involving α -synuclein. However, it should be noted that the samples analysed in this study were obtained from people who had already been diagnosed, and it remains unclear whether the approach could be used as a predictive tool to detect disease at earlier stages. Moreover, it is possible that PMCA is affected by the medication given to the participants who had Parkinson's disease. These people typically receive the hormone dopamine (L-dopa), which has been shown to affect α -synuclein aggregation *in vitro*⁵.

Second, the study adds to a growing body of evidence supporting the 'one polymorph, one disease' hypothesis^{6–8}, which states that different structural forms (polymorphs) of the same aggregated protein can cause distinct pathologies and symptoms. What might lead a protein to adopt different structures? *In vitro*, distinct fold structures can result

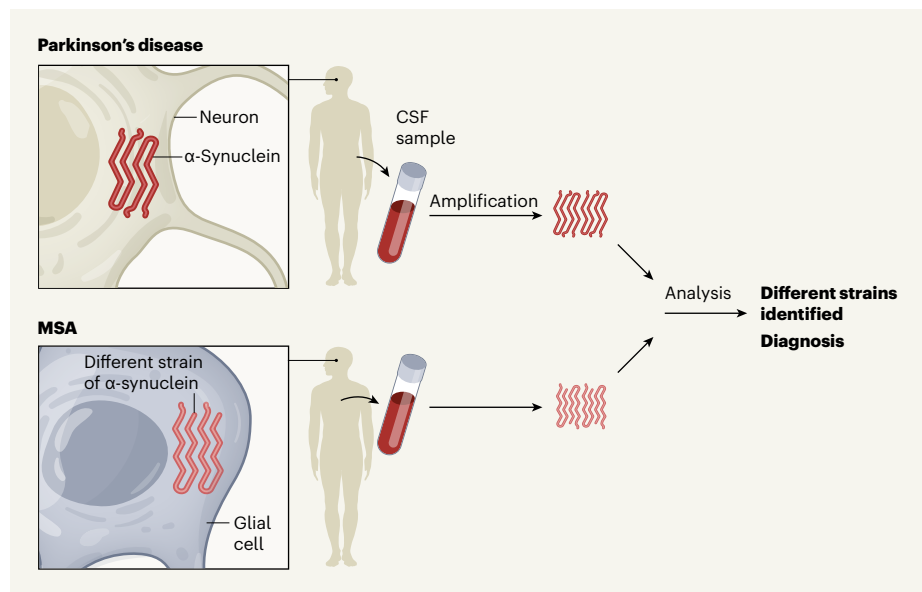


Figure 1 | Different structures for the α -synuclein protein. Two neurodegenerative disorders, Parkinson's disease and multiple system atrophy (MSA), involve aggregates of α -synuclein, which are found in neurons and neuron-supporting glial cells, respectively. Shahnawaz *et al.*¹ have demonstrated that α -synuclein adopts different structures in each disease, indicating that the structure of the protein might contribute to the distinct nature of each disorder. The group extracted tiny amounts of α -synuclein from cerebrospinal fluid (CSF) samples. Protein amplification and analyses revealed different structures for the two samples. These analyses were sufficient to discriminate between the diseases in around 95% of the 200 people studied.

from distinct environmental conditions. For example, different α -synuclein polymorphs arise depending on whether the protein is kept in a phosphate-containing or phosphate-free buffer⁹. *In vivo*, α -synuclein is exposed to several environments. Indeed, the neurons that degenerate in Parkinson's disease and the glia affected in MSA belong to different cell lineages, and have markedly different intracellular environments. In addition, α -synuclein can move between cells, exposing it to both intra- and extracellular environments².

The idea of different polymorphs in disease dates back to studies of prion proteins⁶ in the 1990s. Much like amyloids, prions aggregate in harmful infectious clumps to cause neurodegenerative conditions such as Creutzfeldt-Jakob disease in humans and scrapie in sheep. Several strains of prion, each adopting a different polymorph, typically coexist in a given sample or organism⁷. The strains have different fitnesses in different environments, which governs their ability to replicate⁷ – a phenomenon known as the prion cloud¹⁰.

A corollary of this idea is that if environmental conditions change, the relative abundance of each polymorph might change. This principle also governs the PMCA assay. Under given conditions, the fittest polymorphs should be amplified from a possible mix of pre-existing strains. Indeed, in Shahnawaz and colleagues' experiments, a single distinct polymorph was amplified from Parkinson's disease samples and another from MSA samples.

By contrast, in another recent study that

used PMCA, Strohäker and colleagues¹¹ reported no significant differences between structures of α -synuclein derived from the brains of people who had Parkinson's disease and those with from people with MSA. A possible explanation for this apparent discrepancy is that the two groups used different PMCA

protocols. In addition, Strohäker *et al.* used a much smaller group of patients than did Shahnawaz and colleagues. In fact, analysis using nuclear magnetic resonance spectroscopy did indicate distinct structural features in a subset of Strohäker and colleagues' samples.

High-resolution cryo-electron microscopy has been used to demonstrate the existence of distinct disease-specific polymorphs of another neurodegeneration-associated protein, tau, at atomic resolution⁸. A similar approach using samples extracted under mild conditions might give us a clearer picture of the reality for α -synuclein. Taken together with similar observations for Alzheimer's disease¹², our understanding of the structural landscape of amyloid diseases is broadening.

Juan Atilio Gerez and **Roland Riek** are in the Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, 8093 Zurich, Switzerland. e-mails: juan.gerez@phys.chem.ethz.ch; roland.riek@phys.chem.ethz.ch

1. Shahnawaz, M. *et al.* *Nature* **578**, 273–277 (2020).
2. Peng, C. *et al.* *Nature* **557**, 558–563 (2018).
3. Bousset, L. *et al.* *Nature Commun.* **4**, 2575 (2013).
4. Lau, A. *et al.* *Nature Neurosci.* **23**, 21–31 (2019).
5. Li, J., Zhu, M., Manning-Bog, A. B., Di Monte, D. A. & Fink, A. L. *FASEB J.* **18**, 962–964 (2004).
6. Prusiner, S. B. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
7. Weissmann, C. *PLoS Pathog.* **8**, e1002582 (2012).
8. Zhang, W. *et al.* *eLife* **8**, e43584 (2019).
9. Guerrero-Ferreira, R. *et al.* *eLife* **8**, e48907 (2019).
10. Collinge, J. *Nature* **539**, 217–226 (2016).
11. Strohäker, T. *et al.* *Nature Commun.* **10**, 5535 (2019).
12. Lu, J.-X. *et al.* *Cell* **154**, 1257–1268 (2013).

This article was published online on 5 February 2020.

Medical research

Smoke signals in the DNA of normal lung cells

Gerd P. Pfeifer

Healthy cells in smokers' lungs have a high burden of mutations, similar to the mutational profile of lung cancer. Surprisingly, ex-smokers' lungs have a large fraction of healthy cells with nearly normal profiles. **See p.266**

According to the World Health Organization, there are 1.1 billion smokers worldwide and an estimated 1.8 million deaths from lung cancer annually. Lung cancer caused by smoking can take decades to arise, and smokers have up to a 30-fold higher risk of developing the disease than do non-smokers¹. Carcinogenic components of tobacco smoke promote lung cancer by causing DNA damage that can lead to mutations through known mechanisms,

but what the initial consequences of smoking are for healthy lung cells is poorly understood. On page 266, Yoshida *et al.*² report the mutational profiles of 632 healthy lung cells obtained from whole-genome sequencing of biopsied tissue from 16 individuals: children, adults, non-smokers, current smokers and ex-smokers. The authors analysed the frequency and properties of the mutations present, how they differed according to age

and smoking status, and how these mutations related to those found in a type of lung cancer called squamous-cell carcinoma.

The authors dissociated cells from lung tissue (Fig. 1) and isolated a type of epithelial cell called a basal cell (which can self-renew). Growing single cells into cellular colonies allowed the authors to determine the DNA sequence of the given original cell. A potential caveat of the study is that, although the authors obtained the genome sequences of hundreds of single cells, the number of individuals with each different smoking status was relatively small. The authors report that the number of single nucleotide (point) mutations increased with age – for each extra year of life, about 22 additional such mutations were found per cell.

However, being a former smoker added another 2,330, and being a current smoker added 5,300 point mutations per cell on average, confirming the mutational potency of smoking. Smokers' genomes also had extensive examples of other types of alteration, such as insertion or deletion mutations. The number of mutations in different cells from the same individual could vary by tenfold in smokers, a much higher variability than was found in non-smokers. The stage of the cell cycle at which a cell is exposed to carcinogenic agents might affect how effectively DNA damage is repaired before DNA replication, which could offer an explanation for this high variability.

Yoshida and colleagues examined the mutations in individual cells using previously developed algorithms to focus on all the types of sequence alteration possible (for example, mutation of the DNA base adenine to cytosine, guanine or thymine) and also to assess the bases on either side of a mutated base. Such analysis identifies specific patterns (mutational signatures) that have been used before to characterize the genomes of tumour cells³.

The authors report that the presence of certain mutational signatures increased with age and did not seem to be affected by smoking. These included a signature attributed to natural processes whereby the loss of an amino group in a modified cytosine (termed 5-methylcytosine) changes the base to a thymine. The most common mutational signature in all the samples was one that is rich in cytosine-to-thymine and thymine-to-cytosine mutations. The presence of this signature increased with age and was more common in people with a history of smoking. The underlying processes driving these mutations are unknown. The most common smoking-dependent signature consisted of guanine-to-thymine mutations, a signature that is characteristic of most smoking-associated lung cancers^{4–7}.

Lung cancers have some of the highest mutation frequencies of all tumour types⁸;

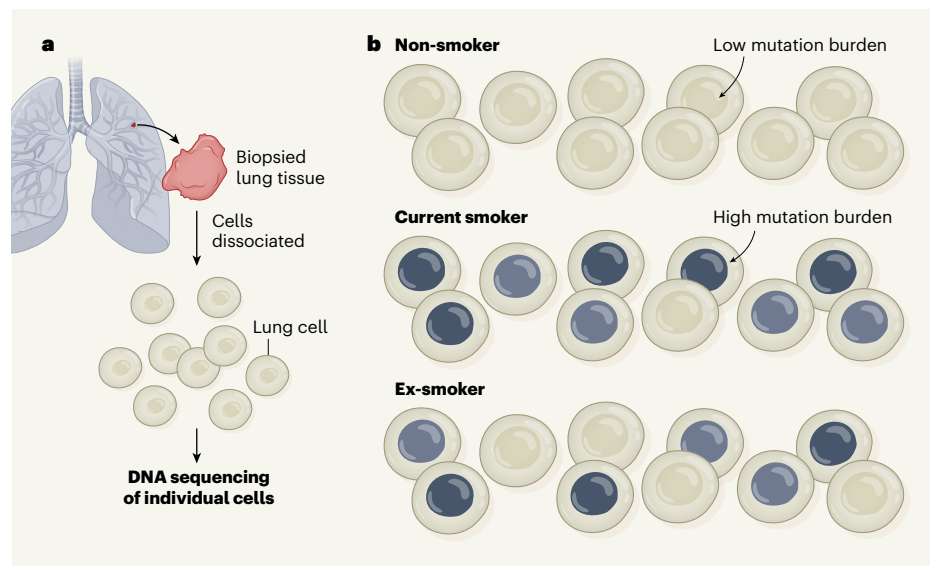


Figure 1 | Mutational burdens in normal human lung cells. Yoshida *et al.*² analysed the pattern of mutations in healthy lung tissue in non-smokers, current smokers and ex-smokers. **a**, Using biopsied lung tissue, the authors determined whole-genome sequences corresponding to single cells. **b**, The cells of the non-smoking individuals had few mutations. By contrast, current smokers had a high proportion of cells with a large number of mutations (grey; darker colour indicates more mutations), and many of these mutations were of a type predominantly found in smokers. Compared with non-smokers, smokers also had greater variability in the mutational load between the different cells of a given individual. Surprisingly, the authors found that five out of six ex-smokers had a substantial fraction (20–50%) of cells that had low numbers of mutations and had hardly any smoking-associated mutational signatures. How these cells arise is a mystery – Yoshida *et al.* speculate that they are generated from a population of as-yet-unknown stem cells.

however, it is thought that only a small number of tumour-promoting (driver) mutations need to occur in a single cell to kick off malignant growth. Given the high mutational burden and the specific smoking-associated mutational signatures found in smokers' healthy epithelial cells, Yoshida and colleagues examined whether these mutations affected crucial genes that are relevant for cancer growth.

Indeed, they found cells that had acquired mutations in genes, including *TP53* and *NOTCH1*, that are driver mutations in squamous-cell carcinomas. These driver mutations were more common in the lung cells of smokers than in those of non-smokers. Some cells even had as many as three driver mutations. However, we do not know how many of these mutations (and in what combination) are required for human lung cancer to develop. Specific *TP53* mutations were found in multiple cells from the same individual, suggesting that these mutations occur early, that cells with the mutation proliferate, or both – similar to what has been observed for sun-exposed healthy human skin⁹.

The higher risk of lung cancer in ex-smokers compared with non-smokers is reflected in their high mutation burden and the signature of smoking-associated mutations in most of their lung cells (similar to the cellular profile of current smokers). Although ex-smokers have a high risk of developing lung cancer, their risk is reduced compared with that of current smokers, and this lowering

depends on the length of time of smoking cessation¹. Why this is the case has been hard to explain. However, perhaps the most surprising result of Yoshida and colleagues' work might offer a clue: in 5 out of 6 ex-smokers, 20–50% of the cells had a low mutation burden that was similar to the profile of non-smokers of the same age range (Fig. 1).

These near-normal cells in ex-smokers had a low frequency of smoking-dependent mutational signatures. Moreover, compared with the ex-smokers' highly mutated cells, these near-normal cells had longer versions of DNA structures called telomeres, which are found at the ends of chromosomes. Telomere length shortens with each cell division; thus, long telomeres suggest that these cells had not undergone many divisions. The authors speculate that these cells might have arisen comparatively recently from divisions of proposed previously dormant (quiescent) stem cells. However, whether such cells exist in human lungs is unknown.

DNA damage can generate a mutation during DNA replication. Therefore, if a population of non-dividing stem cells exists in the human lung, even if exposed to carcinogenic agents, perhaps such cells might avoid incurring mutations if DNA damage is eventually repaired in the absence of division. But the lack of knowledge about these proposed long-lived stem cells and information about the longevity of the different cell types in the human lung make it difficult to explain what occurred in

these ex-smokers' cells with few mutations.

Why do ex-smokers still have a substantial fraction of highly mutated cells that can proliferate, at least when grown *in vitro*? Any short-lived cells that were exposed to carcinogens during their proliferation should have vanished many years after the cessation of smoking. This raises the question of whether there are long-lived differentiated cells in the lung that carry a high mutational burden, and whether these cells can resume proliferation, perhaps because of the plasticity (the ability to change cellular identity) of lung cells¹⁰. A future challenge will be to understand the cell biology of the mechanisms underlying these observations. Perhaps one day it will be possible to develop ways to boost the population of lung cells with few mutations in ex-smokers.

Yoshida and colleagues' study has broadened our understanding of the effects of tobacco smoke on normal epithelial cells in

the human lung. It has shed light on how the protective effect of smoking cessation plays out at the molecular level in human lung tissue and raises many interesting questions worthy of future investigation.

Gerd P. Pfeifer is at the Center for Epigenetics, Van Andel Institute, Grand Rapids, Michigan 49503, USA.
e-mail: gerd.pfeifer@vai.org

1. Lee, P. N., Forey, B. A. & Coombs, K. J. *BMC Cancer* **12**, 385 (2012).
2. Yoshida, K. et al. *Nature* **578**, 266–272 (2020).
3. Alexandrov, L. B. et al. *Nature* **500**, 415–421 (2013).
4. Alexandrov, L. B. et al. *Science* **354**, 618–622 (2016).
5. Pfeifer, G. P. et al. *Oncogene* **21**, 7435–7451 (2002).
6. Denissenko, M. F., Pao, A., Tang, M. & Pfeifer, G. P. *Science* **274**, 430–432 (1996).
7. Kucab, J. E. et al. *Cell* **177**, 821–836 (2019).
8. Kandoth, C. et al. *Nature* **502**, 333–339 (2013).
9. Martincorena, I. et al. *Science* **348**, 880–886 (2015).
10. Tata, P. R. & Rajagopal, J. *Development* **144**, 755–766 (2017).

This article was published online on 29 January 2020.

Forum: Mental health

Digital technology under scrutiny

Does time spent using digital technology and social media have an adverse effect on mental health, especially that of adolescents? Here, two scientists discuss the question, and how digital devices might be used to improve well-being.

The topic in brief

- There is an ongoing debate about whether social media and the use of digital devices are detrimental to mental health.
- Adolescents tend to be heavy users of these devices, and especially of social media.
- Rates of teenage depression began to rise around 2012, when adolescent use of social media became common (Fig. 1).
- Some evidence indicates that frequent users of social media have higher rates of depression and anxiety than do light users.
- But perhaps digital devices could provide a way of gathering data about mental health in a systematic way, and make interventions more timely.

Jonathan Haidt A guilty verdict

A sudden increase in the rates of depression, anxiety and self-harm was seen in adolescents – particularly girls – in the United States and the United Kingdom around 2012 or 2013 (see go.nature.com/2up38hw). Only one suspect was in the right place at the right time to account for this sudden change: social media.

Its use by teenagers increased most quickly between 2009 and 2011, by which point two-thirds of 15–17-year-olds were using it on a daily basis¹. Some researchers defend social media, arguing that there is only circumstantial evidence for its role in mental-health problems^{2,3}. And, indeed, several studies^{2,3} show that there is only a small correlation between time spent on screens and bad mental-health outcomes. However, I present three arguments against this defence.

First, the papers that report small or null effects usually focus on 'screen time', but it is not films or video chats with friends that damage mental health. When research papers allow us to zoom in on social media, rather than looking at screen time as a whole, the correlations with depression are larger, and they are larger still when we look specifically at girls (go.nature.com/2u74der). The sex difference is robust, and there are several likely causes for it. Girls use social media much more than do boys (who, in turn, spend more of their time gaming). And, for girls more than boys, social life and status tend to revolve around intimacy and inclusion versus exclusion⁴, making them more vulnerable to both the 'fear of missing out' and the relational aggression that social media facilitates.

Second, although correlational studies can provide only circumstantial evidence, most of the experiments published in recent years have found evidence of causation (go.nature.com/2u74der). In these studies, people are randomly assigned to groups that are asked to continue using social media or to reduce their use substantially. After a few weeks, people who reduce their use generally report an improvement in mood or a reduction in loneliness or symptoms of depression.

Third, many researchers seem to be thinking about social media as if it were sugar: safe in small to moderate quantities, and harmful only if teenagers consume large quantities. But, unlike sugar, social media does not act just on those who consume it. It has radically transformed the nature of peer relationships, family relationships and daily activities⁵. When most of the 11-year-olds in a class are on Instagram (as was the case in my son's school), there can be pervasive effects on everyone. Children who opt out can find themselves isolated. A simple dose–response model cannot capture the full effects of social media, yet nearly all of the debate among researchers so far has been over the size of the dose–response effect. To cite just one suggestive finding of what lies beyond that model: network effects for depression and anxiety are large, and bad mental health spreads more contagiously between women than between men⁶.

In conclusion, digital media in general undoubtedly has many beneficial uses, including the treatment of mental illness. But if you focus on social media, you'll find stronger evidence of harm, and less exculpatory evidence, especially for its millions of under-age users.

What should we do while researchers hash out the meaning of these conflicting findings? I would urge a focus on middle schools (roughly 11–13-year-olds in the United States), both for researchers and policymakers. Any US state could quickly conduct an informative experiment beginning this September: randomly assign a portion of school districts to ban smartphone access for students in

middle school, while strongly encouraging parents to prevent their children from opening social-media accounts until they begin high school (at around 14). Within 2 years, we would know whether the policy reversed the otherwise steady rise of mental-health problems among middle-school students, and whether it also improved classroom dynamics (as rated by teachers) and test scores. Such system-wide and cross-school interventions would be an excellent way to study the emergent effects of social media on the social lives and mental health of today's adolescents.

Jonathan Haidt is at New York University Stern School of Business, New York, New York 10012, USA.

Nick Allen

Use digital technology to our advantage

It is appealing to condemn social media out of hand on the basis of the – generally rather poor-quality and inconsistent – evidence suggesting that its use is associated with mental-health problems⁷. But focusing only on its potential harmful effects is comparable to proposing that the only question to ask about cars is whether people can die driving them. The harmful effects might be real, but they don't tell the full story. The task of research should be to understand what patterns of digital-device and social-media use can lead to beneficial versus harmful effects⁷, and to inform evidence-based approaches to policy, education and regulation.

Long-standing problems have hampered our efforts to improve access to, and the quality of, mental-health services and support. Digital technology has the potential to address some of these challenges. For instance, consider the challenges associated with collecting data on human behaviour. Assessment in mental-health care and research relies almost exclusively on self-reporting, but the resulting data are subjective and burdensome to collect. As a result, assessments are conducted so infrequently that they do not provide insights into the temporal dynamics of symptoms, which can be crucial for both diagnosis and treatment planning.

By contrast, mobile phones and other Internet-connected devices provide an opportunity to continuously collect objective information on behaviour in the context of people's real lives, generating a rich data set that can provide insight into the extent and timing of mental-health needs in individuals^{8,9}. By building apps that can track our digital exhaust (the data generated by our everyday digital lives, including our social-media use), we

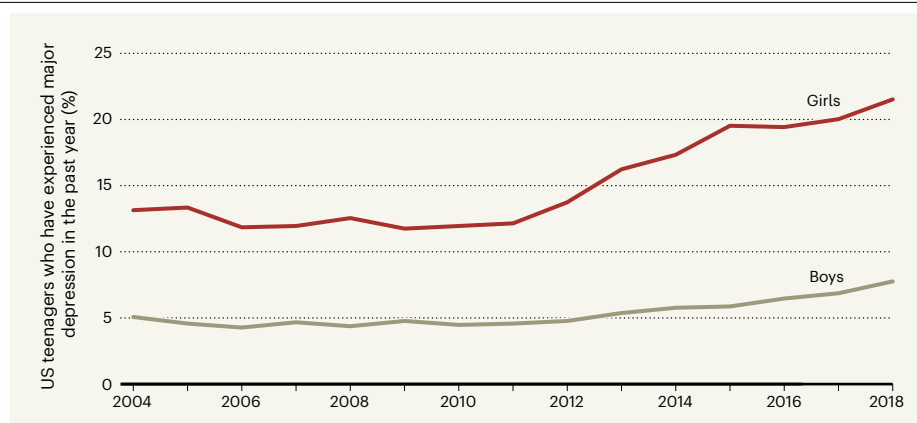


Figure 1 | Depression on the rise. Rates of depression among teenagers in the United States have increased steadily since 2012. Rates are higher and are increasing more rapidly for girls than for boys. Some researchers think that social media is the cause of this increase, whereas others see social media as a way of tackling it. (Data taken from the US National Survey on Drug Use and Health, Table 11.2b; go.nature.com/3ayjaww)

can gain insights into aspects of behaviour that are well-established building blocks of mental health and illness, such as mood, social communication, sleep and physical activity.

These data can, in turn, be used to empower individuals, by giving them actionable insights into patterns of behaviour that might otherwise have remained unseen. For example, subtle shifts in patterns of sleep or social communication can provide early warning signs of deteriorating mental health. Data on these patterns can be used to alert people to the need for self-management before the patterns – and the associated symptoms – become more severe. Individuals can also choose to share these data with health professionals or researchers. For instance, in the Our Data Helps initiative (<https://ourdatahelps.org>), individuals who have experienced a suicidal crisis, or the relatives of those who have died by suicide, can donate their digital data to research into suicide risk.

Because mobile devices are ever-present in people's lives, they offer an opportunity to provide interventions that are timely, personalized and scalable. Currently, mental-health services are mainly provided through a century-old model in which they are made available at times chosen by the mental-health practitioner, rather than at the person's time of greatest need. But Internet-connected devices are facilitating the development of a wave of 'just-in-time' interventions¹⁰ for mental-health care and support.

A compelling example of these interventions involves short-term risk for suicide^{9,11} – for which early detection could save many lives. Most of the effective approaches to suicide prevention work by interrupting suicidal actions and supporting alternative methods of coping at the moment of greatest risk. If these moments can be detected in an individual's digital exhaust, a wide range of intervention options become available, from providing information about coping skills

and social support, to the initiation of crisis responses. So far, just-in-time approaches have been applied mainly to behaviours such as eating or substance abuse⁸. But with the development of an appropriate research base, these approaches have the potential to provide a major advance in our ability to respond to, and prevent, mental-health crises.

These advantages are particularly relevant to teenagers. Because of their extensive use of digital devices, adolescents are especially vulnerable to the devices' risks and burdens. And, given the increases in mental-health problems in this age group, teens would also benefit most from improvements in mental-health prevention and treatment. If we use the social and data-gathering functions of Internet-connected devices in the right ways, we might achieve breakthroughs in our ability to improve mental health and well-being.

Nick Allen is at the Center for Digital Mental Health, University of Oregon, Eugene, Oregon 97403-1227, USA.
e-mail: nallen3@uoregon.edu

- Twenge, J. M., Martin, G. N. & Spitzberg, B. H. *Psychol. Pop. Media Culture* **8**, 329–345 (2019).
- Orben, A. & Przybylski, A. K. *Nature Hum. Behav.* **3**, 173–182 (2019).
- Odgers, C. L. & Jensen, M. R. *J. Child Psychol. Psychiatry* <https://doi.org/10.1111/jcpp.13190> (2020).
- Maccoby, E. E. *The Two Sexes: Growing Up Apart, Coming Together* Ch. 2 (Harvard Univ. Press, 1999).
- Nesi, J., Choukas-Bradley, S. & Prinstein, M. J. *Clin. Child. Fam. Psychol. Rev.* **21**, 267–294 (2018).
- Rosenquist, J. N., Fowler, J. H. & Christakis, N. A. *Molec. Psychiatry* **16**, 273–281 (2011).
- Orben, A. *Social Psychiatry Psychiatr. Epidemiol.* <https://doi.org/10.1007/s00127-019-01825-4> (2020).
- Mohr, D. C., Zhang, M. & Schueller, S. M. *Annu. Rev. Clin. Psychol.* **13**, 23–47 (2017).
- Nelson, B. W. & Allen, N. B. *Perspect. Psychol. Sci.* **13**, 718–733 (2018).
- Nahum-Shani, I. et al. *Ann. Behav. Med.* **52**, 446–462 (2018).
- Allen, N. B., Nelson, B. W., Brent, D. & Auerbach, R. P. *J. Affect. Disord.* **250**, 163–169 (2019).

N.A. declares competing financial interests: see go.nature.com/2oq2riu for details.

This article was published online on 10 February 2020.

The promise and challenge of therapeutic genome editing

<https://doi.org/10.1038/s41586-020-1978-5>

Jennifer A. Doudna^{1,2,3,4,5,6,7*}

Received: 10 February 2019

Accepted: 20 November 2019

Published online: 12 February 2020

Genome editing, which involves the precise manipulation of cellular DNA sequences to alter cell fates and organism traits, has the potential to both improve our understanding of human genetics and cure genetic disease. Here I discuss the scientific, technical and ethical aspects of using CRISPR (clustered regularly interspaced short palindromic repeats) technology for therapeutic applications in humans, focusing on specific examples that highlight both opportunities and challenges. Genome editing is—or will soon be—in the clinic for several diseases, with more applications under development. The rapid pace of the field demands active efforts to ensure that this breakthrough technology is used responsibly to treat, cure and prevent genetic disease.

In the nearly seventy years since the discovery of the DNA double helix, technologies have advanced for the determination, analysis and alteration of genome sequences and gene-expression patterns in cells and organisms. These molecular tools are the foundation of molecular biology, driving the therapeutic industry by increasing the understanding of the genetics of normal and disease traits. The ability to diagnose genetic diseases has developed rapidly with reductions in the costs of genome sequencing, increases in comparative analyses of human genome sequences and increased applications of high-throughput genomic screening. However, the dearth of therapies, much less cures, for genetic diseases has created a growing separation between diagnostics and treatments, underscoring the urgent need to develop therapeutic options. Mitigation or correction of disease-causing mutations is a tantalizing goal with tremendous potential to save and improve lives, representing a convergence of technical and medical advances that could eventually eradicate many genetic diseases.

Although methods for genome engineering and gene therapy have been of interest for decades, the development of engineered and programmable enzymes for the manipulation of DNA sequences has driven a biotechnological revolution^{1–5}. In particular, fundamental research showing how CRISPR and CRISPR-associated (Cas) proteins provide microorganisms with adaptive immunity has propelled transformative technological opportunities enabled by RNA-guided proteins. CRISPR–Cas9 and related enzymes have been used to manipulate the genomes of cultured and primary cells, animals and plants, vastly accelerating the pace of fundamental research and enabling breakthroughs in agriculture and synthetic biology^{6–9}. Building on past gene therapy efforts¹⁰, we are entering an era in which genome-editing tools will be used to inactivate or correct disease-causing genes in patients, offering life-saving cures to people who have genetic disorders.

In this Review, I discuss the therapeutic opportunities of genome editing, the ability to alter the DNA in cells and tissues in a site-specific manner. In addition to presenting current capabilities and limitations of the technology, I also describe what it will take to apply therapeutic genome editing in the real world. A comparison of somatic-cell and

germline editing highlights the importance of open public discussion about, and regulation of, this powerful technology.

The scope of genome-editing applications

Although the genetics of human disease are often complex, some of the most common genetic disorders stem from mutations in a single gene. Cystic fibrosis, Huntington's chorea, Duchenne muscular dystrophy (DMD) and sickle cell anaemia each represent diseases that result from defects in only one gene in the human genome; such monogenic diseases, of which more than 5,000 are known, affect at least 250 million individuals globally. DNA sequencing of affected families has provided detailed information about the mutations that lead to each disorder, as well as correlations between specific genetic changes (genotype) and disease severity. These data in turn reveal DNA sequence alterations or corrections that could provide a genetic cure by either disrupting the function of a toxic or inhibitory gene or restoring the function of an essential gene.

Sickle cell disease and muscular dystrophy, two common human genetic disorders, provide instructive examples of diseases that could be treated or cured by genome editing in the foreseeable future. Sickle cell disease results from a single base-pair change in the DNA that in turn generates a defective protein with destructive consequences in red blood cells. DMD belongs to a set of muscle-wasting diseases that result from DNA sequence changes that disrupt the normal production of a protein required for muscle strength and stability. A closer look at each of these diseases illustrates the ways that genome editing could offer therapeutic benefit to patients.

Sickle cell disease occurs in individuals who have two defective copies of the gene that encodes β -globin (*HBB*), the protein required to form oxygen-carrying haemoglobin in adult blood cells. Described originally by Linus Pauling and colleagues¹¹ and mapped to a genetic locus in the 1950s¹², a single A-to-T mutation results in a glutamate-to-valine substitution in β -globin (Fig. 1). This seemingly small change causes the defective protein to form chain-like polymers of haemoglobin, inducing red blood cells to assume a sickled shape that leads to occluded blood

¹Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, USA. ²Department of Chemistry, University of California Berkeley, Berkeley, CA, USA. ³California Institute for Quantitative Biosciences (QB3), University of California Berkeley, Berkeley, CA, USA. ⁴Innovative Genomics Institute, University of California Berkeley, Berkeley, CA, USA. ⁵Howard Hughes Medical Institute, University of California Berkeley, Berkeley, CA, USA. ⁶MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁷Gladstone Institutes, University of California San Francisco, San Francisco, CA, USA. *e-mail: doudna@berkeley.edu

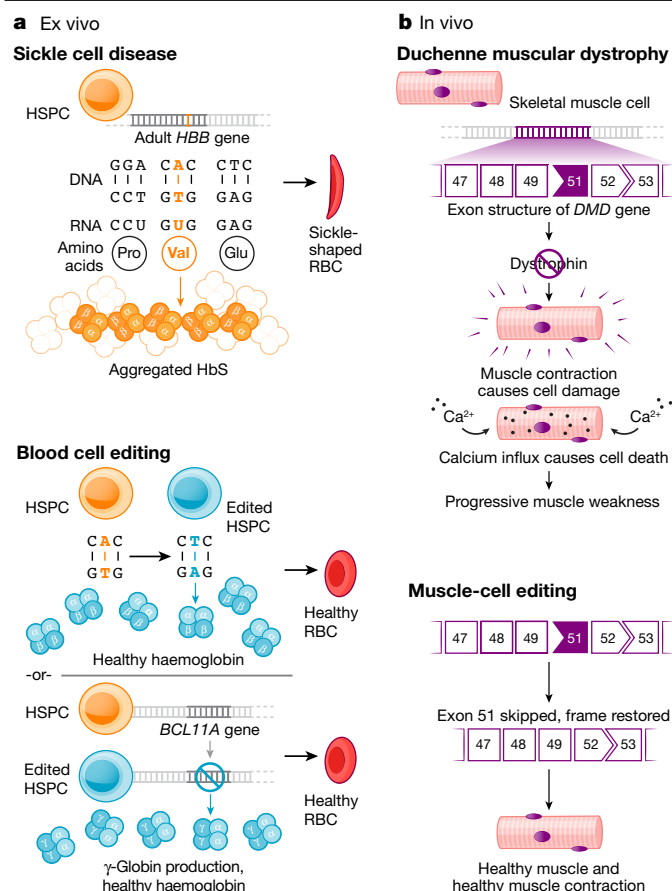


Fig. 1 | Ex vivo and in vivo genome editing to treat human disease.

a, b, Somatic genome-editing treatments may be accomplished in one of two ways: by removing and editing target cells in the laboratory before returning them to the patient (ex vivo, **a**) or by directly delivering CRISPR–Cas editing tools to the affected tissue (in vivo, **b**). **a**, Blood disorders such as sickle cell disease may be treated by editing haematopoietic stem or progenitor cells (HSPCs) ex vivo, creating normal red blood cells (RBCs). **b**, Disorders that affect non-removable tissues, such as DMD, require editing of affected cell types (in this case myogenic cells) in vivo.

vessels, pain and life-threatening organ failure. Although bone-marrow transplantation can cure the disease, it requires the use of cells from an individual whose immune profile matches that of the patient. In principle, sickle cell disease could be cured by removing blood stem cells—that is, haematopoietic progenitor cells—from a patient and using genome editing to either correct the disease-causing mutation in β -globin or activate expression of γ -globin, a fetal form of haemoglobin that could substitute for defective β -globin (Fig. 1). The edited stem cells could then be transplanted back into the patient, in whom the progeny of these edited stem cells would produce healthy red blood cells.

The ability to edit cells extracted from patients with sickle cell disease makes this disease—and other blood disorders—one of the more tractable pathologies that could be treated by genome editing in the near future. Most genetic diseases, however, will require genome editing of cells in the body (in situ) to correct a genetic defect associated with a disease. Muscular dystrophy exemplifies this type of disorder, because it involves the weakening and disruption of skeletal muscles over time^{13,14}. The most common type, DMD, affects 1 in 5,000 males at birth, who inherit mutations in the gene that encodes dystrophin (*DMD*), a scaffolding protein that maintains the integrity of striated muscles (Fig. 1). Over time, these patients lose the ability to walk and eventually succumb to respiratory and heart failure, typically dying by the third decade of life. In contrast to therapies that delay disease

progression, genome editing offers the possibility of permanent restoration of the missing dystrophin protein. Although more than 3,000 different mutations can cause DMD, most occur at hotspots within *DMD*. Notably, restoration of a small percentage (around 15%) of the normal expression levels of dystrophin can provide a clinical benefit¹⁵.

To treat or cure monogenetic disorders such as sickle cell disease and DMD, it will be important to match the underlying genetic defect with the best genome-editing approach. In each case, this involves multiple considerations, including the type of editing needed, the mode of cell or tissue delivery required and the extent of gene knockout or correction that will provide therapeutic value.

The next section describes current genome-editing technologies that offer the potential of curative human genome editing.

Genome-editing strategies

Engineered DNA-cleaving enzymes, including zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), have demonstrated the potential of therapeutic genome editing. These early technologies enabled the inactivation of the gene encoding the HIV co-receptor CCR5 in somatic cells¹⁶, mitigation of the *HBB* gene mutation in haematopoietic stem cells^{17,18} and engineering of immune cells for the treatment of childhood cancer¹⁹. To realize this potential, the development of CRISPR–Cas9 for genome editing offers a simpler technology that has been adopted widely owing to the ease of programming of its DNA-binding and modifying capabilities. Cas9 is a protein that assembles with a guide RNA—either as separate crRNA and tracrRNA components or a chimeric single-guide RNA (sgRNA)—to create a molecular entity that is capable of binding and cutting DNA¹. Notably, DNA binding occurs at a 20-base-pair DNA sequence that is complementary to a 20-nucleotide sequence in the guide RNA and that can be readily altered by the researcher^{1,20} (Fig. 2). The DNA-recognition site must be adjacent to a short motif (the protospacer adjacent motif or PAM) that acts as a switch, triggering Cas9 to make a double-stranded DNA break within the target sequence^{1,20}. In cells of all multicellular organisms, including humans, such double-stranded DNA breaks induce DNA repair by endogenous cellular pathways that can introduce alterations to the DNA sequence, including small sequence changes or genetic insertions^{21,22}. Although CRISPR–Cas9-induced genome editing is effective in almost all cell types, controlling the exact editing outcome remains a challenge in the field, as discussed below.

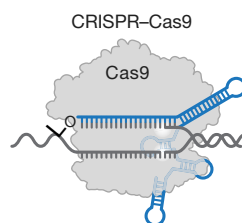
Although the Cas9 of *Streptococcus pyogenes* (SpCas9) is the enzyme that is most commonly used for genome editing and genetic manipulation using CRISPR–Cas, a growing collection of natural and engineered Cas9 homologues and other CRISPR–Cas RNA-guided enzymes is expanding the genome-manipulation toolbox^{6,23,24}. It is the intrinsic programmability that is present in this diversity of enzymes that underscores the utility of CRISPR–Cas technology for genome editing and other applications including gene regulation and diagnostics (Fig. 2).

For safe and effective clinical use ex vivo and in vivo, genome editing needs to be accurate, efficient and deliverable to the desired cells or tissues. CRISPR–Cas9-generated DNA cleavage induces genome editing during double-stranded DNA break repair by non-homologous end joining and/or homology-directed repair (Fig. 2). Homology-directed repair, which requires the presence of a DNA template, is—in most cases—used by the cell less frequently than non-homologous end joining. Furthermore, both types of repair can happen in the same cell, creating different alleles of an edited gene. Two concurrent double-stranded DNA breaks can induce chromosomal translocations. For these reasons, an active area of CRISPR–Cas technology development involves controlling DNA repair outcomes to ensure that the desired genetic change is introduced.

Alternatives to DNA-cleavage-induced editing include using CRISPR–Cas9 to directly alter the chemical sequence (base editing)^{25,26}, to generate RNA templates for gene alteration (prime editing)^{27,28} and for transcriptional control (CRISPR interference and

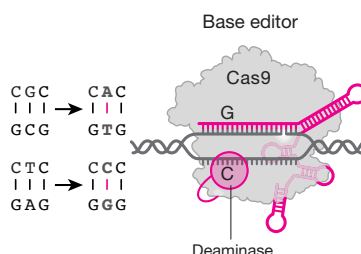
a Genome editing

- Insertion of gene(s), deletion of gene(s), replacement of gene(s) by DSB
- 1–1,000s of nucleotides
- Permanent
- Other tools: meganucleases, TALENs, ZFNs, other CRISPR nucleases



b Base editing

- Single-bp change by DNA nick
- SNP reversal; gene KO
- Permanent



c Gene regulation

- Gene repression
- Temporary or persistent
- Epigenetic modification or RNA targeting
- Gene activation
- Temporary or persistent
- Epigenetic modification

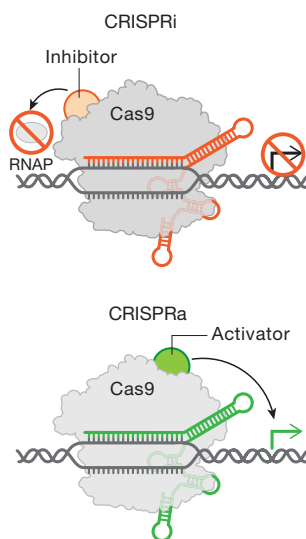


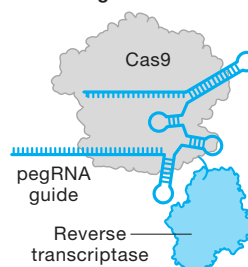
Fig. 2 | The genome editing toolbox. **a–c**, Most well-validated CRISPR-based tools perform one of three functions: genome editing (**a**), base editing (**b**) or gene regulation (**c**). These systems rely on RNA-guided Cas9 or Cas12a to target specific genomic sites. These techniques edit the target site by direct cleavage of one or both nuclease active sites, triggering cellular DNA repair by non-homologous end joining or homology-directed repair, and/or by relying on fused effector proteins. **a**, CRISPR–Cas9 generates a double-stranded break (DSB) at the target site to simulate endogenous DNA repair. These double-stranded breaks are resolved by the endogenous cellular repair machinery, resulting in one of two main outcomes at the cut site: an insertion or deletion, or the insertion of or replacement with donor DNA that is delivered at the same time. **b**, A fused domain replaces a single base through deamination and DNA replication or repair. This single base change is propagated to the complementary strand of DNA. Changes include C to U (uracil), which is swapped to a T during replication or repair, and A to I (inosine), which is treated as a G. bp, base pair; KO, knockout. **c**, CRISPR-mediated gene repression or interference (CRISPRi) sterically blocks the RNA polymerase and induces heterochromatinization, leading to direct epigenetic modifications such as DNA methylations or RNA targeting by modifying individual bases or RNA cleavage. CRISPR-mediated gene activation (CRISPRa) recruits the transcription machinery to increase expression of the target region and leads to direct epigenetic modifications such as histone acetylation.

CRISPR activation)^{29,30} (Fig. 3). In addition, it may be possible to control gene outputs through Cas9-mediated epigenetic modification^{31,32}. Although these methods have been used in cultured cells, they are not

Tool

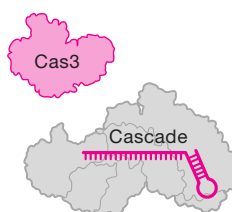
Outcome at target site

a Prime editing



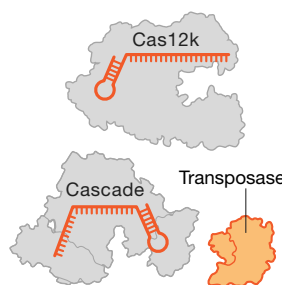
- Insertion, deletion, replacement
- Nicks
- Reverse transcriptase fused to Cas9 nickase with 3' extended pegRNA guide

b CRISPR–Cas3



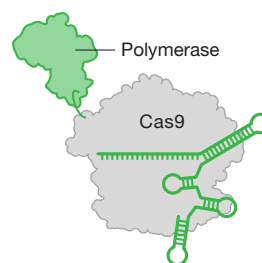
- Large deletion (~0.5–100 kb)
- Multiple cuts
- Naturally occurring crRNA-guided Cascade targeting complex and nuclease-helicase Cas3

c CRISPR-associated transposases



- Large insertion (~0.3–10 kb)
- Naturally occurring Cas12k or Cascade and genomically associated Tn7-like transposases

d EvolvR



- Continuous mutagenesis
- Nicks
- Error-prone, nick-translating DNA polymerase fused to Cas9 nickase

Fig. 3 | Emerging tools. New modifications of the CRISPR–Cas platform are currently underway and, if validated, could provide specific genome modification specialties. **a**, Cas9 binds to and nicks the genomic target, after which the reverse transcriptase copies the sequence of the prime-editing guide RNA (pegRNA) to the target site. **b**, Cascade binds to a genomic target, inducing processive cleavage by Cas3 and generating large deletions. **c**, Cascade or Cas12k binds to the genomic target and directs donor DNA insertion by the Tn7-like transposase. **d**, Cas9 binds to and nicks the genomic target, after which the error-prone polymerase generates diversity in an adjacent window, thus enabling directed evolution.

yet ready for clinical use until matters of specificity^{33,34} and delivery are addressed.

Two strategies to mitigate or cure sickle cell disease take advantage of demonstrated strategies for site-specific genome editing (Figs. 1, 2). The first involves the restoration of the wild-type *HBB* gene sequence by homology-directed repair³⁵. The second approach is to activate expression of γ -globin, the fetal form of haemoglobin that is typically silenced in adult cells, by disrupting γ -globin repressors^{36–41} or their

binding sites in the promoter of the γ -globin (*HBG1/HBG2*) genes^{40,42,43}. These genome-editing strategies require the collection of a patient's haematopoietic stem and progenitor cells, either to correct the mutation in *HBB* or to restart expression of γ -globin, and the subsequent reintroduction of the edited cells into the bone marrow. Major progress in the delivery⁴⁴ and handling of haematopoietic stem and progenitor cells has resulted in impressive efficiencies of mutation correction or mitigation^{18,45–47} that are expected to be curative.

Such an approach, although it requires a bone-marrow transplantation, would remove the need for a compatible bone-marrow donor and thus provide a path for treating and potentially curing many more people than can currently be treated. As discussed below, improvements in in vivo delivery technology may one day enable treatment without requiring bone-marrow transplantation, which would reduce both expense and patient hardship.

Whereas in vivo editing may resolve some of the issues with ex vivo sickle cell therapies, studies in DMD illustrate that other challenges arise when attempting in situ gene correction. Three reports^{48–50} have highlighted both the tremendous potential of genome editing and the considerable challenges that remain before genome editing can be used to treat or cure muscular dystrophy in humans. In the first study, a mouse model of DMD was created using CRISPR–Cas9 to generate a common deletion (Δ Ex50) in the *Dmd* gene that also occurs in patients with DMD⁴⁸. The severe muscle dysfunction in the Δ Ex50 mice was corrected by systemic delivery of an adeno-associated virus (AAV) that encoded the CRISPR–Cas9 genome-editing components, restoring up to 90% of dystrophin protein expression throughout the skeletal muscles and hearts of Δ Ex50 mice. The second study used CRISPR–Cas9-mediated genome editing to remove a mutation in exon 23 in the *mdx* mouse model of DMD, providing partial recovery of functional dystrophin protein in skeletal myofibres and cardiac muscle^{25,26,49}. In the third study, dogs with the Δ Ex50 mutation, which corresponds to a mutational 'hotspot' in the human *DMD* gene, were treated using CRISPR–Cas9⁵⁰. After virus-mediated systemic delivery in skeletal muscle, dystrophin levels were restored to 3–90% of normal, and the appearance of the muscle tissue in treated dogs was improved. Although promising, these reports, as well as early-stage data from patients treated with in vivo gene editing using ZFNs, highlight the gap between animal studies and applications in humans^{51–53} and underscore the need for improved methods for in situ delivery, as discussed in the next section. An early-stage clinical trial in which in vivo CRISPR–Cas9 delivery to the eye is used to treat congenital blindness⁵⁴ and a close-to-the-clinic program for liver gene editing⁵⁵ will soon provide key first-in-human data to inform the direction of that effort.

Towards tissue-specific delivery

For any of these genome-editing methods to be useful clinically, the CRISPR–Cas enzymes, associated guide RNAs and any DNA repair templates must make their way into the cells that are in need of genetic repair. To produce a functional genome-editing complex, Cas9 and sgRNA can be introduced to cells in target organs in formats that include DNA, mRNA and sgRNA, or protein and sgRNA. All three formats are currently—or will soon be—used in the clinic, using viral vectors, nanoparticles and electroporation of protein–RNA complexes, and each has distinct benefits and limitations (Table 1). The currently favoured form of ex vivo delivery to primary cells is electroporation of Cas9 as a preformed protein–RNA (ribonucleoprotein (RNP)) complex^{44,56}. In vivo delivery, which is much more challenging, is currently conducted using viral vectors (typically AAVs) or lipid nanoparticles bearing Cas9 mRNA and an sgRNA. The difficulty of ensuring efficient, targeted delivery into desired cells in the body currently limits the clinical opportunities of in vivo genome editing, although this is an area of increasing research and development.

Viral delivery vehicles, including lentiviruses, adenoviruses and AAVs, offer advantages of efficiency and tissue selectivity (Table 1). AAVs are

attractive because of the reduced risk of genomic integration, inherent tissue tropism and clinically manageable immunogenicity. In addition, long-term expression of trans-genes that encode Cas9 and sgRNA from the episomal viral genome could help to boost genome-editing efficiency in patients, such as individuals with DMD as discussed below⁵⁷. Notably, the FDA has approved the use of AAVs for gene-replacement therapy in patients with spinal muscular atrophy and congenital blindness, and clinical trials are in progress⁵⁸.

There are, however, considerable challenges to using AAVs for the therapeutic delivery of CRISPR–Cas components. First, the AAV genome can only encode around 4.7 kilobases (kb) of genetic cargo, less than other viral vectors and not much larger than the 4.2-kb length of the gene that encodes *S. pyogenes* Cas9. As a result, for applications that necessitate the insertion of a corrective gene, a second AAV vector that encodes the sgRNA and a template sequence for homology-directed DNA repair must be used, reducing efficiency owing to the need for cells to acquire both AAV vectors at once^{59,60}. Smaller genome-editing proteins, such as the Cas9 of *Staphylococcus aureus* or *Campylobacter jejuni* and other newly identified CRISPR–Cas enzymes, may circumvent this issue^{23,61–65}. Second, long-term expression of genome-editing molecules may expose patients to undesired off-target editing or immune reactions^{66,67}. Third, the production of AAVs at scale and the use of good manufacturing process methods at affordable cost for clinical use remain formidable challenges^{68–70}.

Nanoparticles offer an alternative to virus-based delivery of Cas9 and sgRNAs and are suitable for delivering genome-editing components in the form of DNA, mRNA or RNPs (Table 1). For example, the delivery of lipid-mediated nanoparticles has been used to transport CRISPR–Cas components in the form of either mRNA and sgRNA or preassembled RNPs into tissues^{71–74}. When combined with a highly anionic sgRNA, the cationic Cas9 protein forms a stable RNP complex that has anionic properties suitable for encapsulation by cationic lipid nanoparticles, potentially enabling delivery into cells through endocytosis and macropinocytosis. Cationic lipid-based delivery is a relatively easy, low-cost process for delivering CRISPR components into cells⁷⁵. This approach has been used for one-shot delivery of Cas9 RNPs into mice to achieve therapeutically useful levels of genome editing in the liver⁵⁵. Disadvantages of this approach include marked toxicity of the lipid-mediated nanoparticles⁷⁶ and the potentially undesired selectivity of cell-type-specific uptake of the particles.

Inorganic nanoparticles are another type of delivery vehicle with advantages that include tunable size and surface properties. Gold nanoparticles, in particular, are attractive materials for molecular delivery because of the intrinsic affinity of gold for sulfur, enabling functionalized molecules to be coupled to the gold particle surface. Gold nanoparticles were used originally for nucleic acid delivery by conjugating to thiol-linked DNA or RNA⁷⁷. Cas9 protein–sgRNA complexes can be incorporated by assembly with DNA-linked particles⁷⁸. Such assemblies, complexed with polymers capable of disrupting endosomes and including DNA templates for homology-directed repair, were found to promote correction of *Dmd* gene mutations in mice⁷⁹. Ongoing research continues to advance nanoparticle delivery technology, such as for endothelial cells that could enable access to the lungs and other organs⁸⁰.

Strategies for non-viral cellular delivery of CRISPR–Cas components include electroporation, which involves pulsing cells with high-voltage currents that create transient nanometre-sized pores in the cell membrane. This process allows negatively charged DNA or mRNA molecules or CRISPR–Cas RNPs to enter the cells. Although this method is a primary method of Cas9–sgRNA delivery to cells ex vivo, electroporation has also been used successfully for Cas9 delivery to animal zygotes^{81,82}, and to introduce CRISPR–Cas constructs directly into the skeletal muscle in mice, resulting in restoration of *Dmd* gene expression⁸³. Electroporation will likely be of limited utility for most in vivo genome-editing applications because of its impracticality.

Table 1 | Methods for delivering genome-editing tools

Property	Nanoparticles	Viruses	RNPs
Features and applications	Cationic lipid polymers can be used to encapsulate molecular cargo, facilitating cellular entry.	AAVs are the most commonly used clinical delivery vehicle for gene therapy.	Purified protein and guide RNA can be electroporated into stem cells extracted from patients to treat blood disorders such as sickle cell disease.
Size	50–500 nm	20 nm	12 nm
Payload	mRNA, DNA, RNP (from most to least commonly used)	DNA	Preformed enzyme complexes
Advantages	<ul style="list-style-type: none"> - Inexpensive and relatively easy to produce - No genomic integration - Low immunogenicity 	<ul style="list-style-type: none"> - Broad tissue targeting possibilities - Clinically established method - Efficient 	<ul style="list-style-type: none"> - No genomic integration - No long-term expression and fewer off-target effects
Disadvantages	<ul style="list-style-type: none"> - Limited capacity for tissue targeting 	<ul style="list-style-type: none"> - Limited cargo size - Undesired integration risk - Sustained expression can lead to off-target effects - Immunogenicity - High cost and manufacturing challenges 	<ul style="list-style-type: none"> - Will not enter cells without engineering or additional reagents - Potential immunogenicity in vivo - Unprotected RNPs are at risk of degradation
Targets	Liver	Liver, eyes, brain, lungs and muscle	Oocytes, stem cells and T cells

The three main delivery strategies that could be used for clinical genome-editing applications are nanoparticles, viruses and purified RNPs. The approaches vary in important ways, which generally limit their suitability for editing to specific cell or tissue types.

Another non-viral delivery method is the direct application of pre-assembled CRISPR–Cas RNPs, with or without chemical modifications to assist cell penetration of cultured cells or organs. This delivery mode can reduce possible off-target mutations relative to delivering Cas9-encoding DNA or mRNA due to the short half-life of RNPs^{76,84–86}. New strategies for the direct delivery of CRISPR–Cas9 RNP complexes continue to emerge, including those using molecular engineering to enhance the targeting of specific cell types⁸⁷ and to increase the efficiency of cell penetration⁸⁸.

Delivery remains perhaps the biggest bottleneck to somatic-cell genome editing, a reality that has motivated increasing effort across different disciplines. Emerging strategies that may have substantial impact on the clinical use of genome editing include advances in nanoparticle- and cell-based delivery methods⁸⁹ as well as approaches that involve red blood cells⁹⁰ and nanowires⁹¹.

Accuracy, precision and safety of genome editing

The clinical utility of genome editing depends fundamentally on accuracy and precision. Accuracy refers to the ratio of on- versus off-target genetic changes, whereas precision relates to the fraction of on-target edits that produce the desired genetic outcome. Inaccurate (off-target) genome editing occurs when CRISPR-induced DNA cleavage and repair happens at genome locations not intended for modification, typically sites that are close in sequence to the intended editing site⁹². Imprecise genome editing results from different modes of DNA repair after on-target DNA cleavage, such as a mixture of non-homologous end-joining and homology-directed recombination events that produce different sequences at the desired editing location in different cells. In addition, large deletions and complex genomic rearrangements have been observed after genome editing in mouse embryonic cells, haematopoietic progenitor cells and human immortalized epithelial cells^{93–95}. Although these events occur at low frequency, they could be important in a clinical setting if rare translocations led to cancer^{96–98}. Careful testing will be required to detect and monitor both the accuracy and precision of genome editing in clinical settings and ultimately to reduce or eliminate undesired events by controlling target site recognition and DNA repair outcomes. The National Institute of Standards and Technology manages a scientific consortium that aims to measure and standardize such outcomes as genome-editing technology advances⁹⁹.

The risks intrinsic to DNA-cleavage-induced genome editing have spurred the development of CRISPR–Cas9-mediated genome

regulation or editing methods that do not involve double-stranded DNA cutting. CRISPR interference and CRISPR activation both use catalytically deactivated forms of Cas9 (dCas9) that are fused to transcriptional repressors or activators^{29,100}. Similarly, CRISPR–Cas9-mediated epigenetic modification to control gene expression is also under development¹⁰¹. An alternative approach is to use CRISPR–Cas9 coupled to DNA-editing enzymes that catalyse targeted A-to-G or C-to-T genomic sequence changes without inducing a break in the DNA, potentially reversing pathogenic single-nucleotide changes or disabling genes through the introduction of a stop codon^{25,26}. CRISPR–Cas9 can also be linked to reverse transcriptase and used for targeted template-directed sequence alterations¹⁰². All of these strategies—although elegant in principle—involve large chimeric proteins that pose additional challenges of delivery into primary cells or animals. The specificity of action, both at the target site and genome-wide, remains an area of active investigation. Issues of delivery, potency and specificity of CRISPR interference, CRISPR activation and CRISPR-mediated base editing and prime editing will need to be thoroughly addressed before they are ready for clinical use.

Other factors that affect clinical applications of genome editing include the immunogenicity of bacterially derived editing proteins, the potential for pre-existing antibodies against CRISPR components to cause inflammation and the unknown long-term safety and stability of genome-editing outcomes. Immunogenicity of CRISPR–Cas proteins could be managed by high-efficiency one-time editing treatments and by using different editing enzymes. Pre-existing Cas9 antibodies and reactive T cells have been detected in humans exposed to pathogenic bacteria that have CRISPR systems, although it is unknown whether these are present at sufficiently high concentrations to trigger an immune response to the genome-editing enzymes^{66,103}. Notably, genome-editing therapies that involve ex vivo editing, such as for sickle cell disease, are not as affected by either immunogenicity or pre-existing CRISPR–Cas antibodies, as the natural decay of residual Cas9 protein in the ex vivo edited cells minimizes Cas9 exposure. The potential for inadvertent selection of genome-edited cells with undesired genetic changes came to light with the observation that selection for inactivation of the p53 pathway, which is associated with rapid cell growth and cancer, can occur during laboratory experiments on cells that are not used clinically^{104,105}. Subsequent experiments showed that p53 inactivation can be controlled or avoided through protocol optimization^{47,106}. As for the long-term safety and efficacy of genome-edited cells in vivo, much remains to be determined. However, the recent report

Review

of a single HIV-positive patient who received CRISPR–Cas9-edited haematopoietic progenitor cells showed that although the number of edited cells was too low to mitigate HIV infection, no adverse outcome was detected more than 19 months after transplantation of the edited cells¹⁰⁷. Together, these findings suggest that there are, at present, no known insurmountable hurdles to the eventual development of safe and effective clinical applications of genome editing in humans.

Therapeutic genome editing

The clinical potential of genome editing exemplified by applications in sickle cell disease, muscular dystrophy and other monogenetic disorders could be stymied by extreme pricing of such next-generation therapeutics. Although CRISPR technology itself is a democratizing tool for scientists, extension of its broad utility in biomedicine requires addressing the costs of development, personalization for individual patients and the intrinsic difference between a chronic disease treatment versus a one-and-done cure¹⁰².

Current clinical trials using the CRISPR platform aim to improve chimeric antigen receptor (CAR) T cell effectiveness, treat sickle cell disease and other inherited blood disorders, and stop or reverse eye disease¹⁰⁸. In addition, clinical trials to use genome editing for degenerative diseases including for patients with muscular dystrophy are on the horizon. For sickle cell disease, the uniform nature of the underlying genetic defect lends itself to correction by a standardized CRISPR modality that could be used in many if not most patients. This simplifies clinical testing but also makes the need to address patient cost and access more acute, given that the approximately 100,000 US patients and millions of individuals in African and Asian countries will be candidates for treatment.

For muscular dystrophy, the genetic diversity among patients lends itself to personalization, which is an inherent strength of the CRISPR genome-editing platform; however, it also complicates clinical testing strategies. In addition, progressive diseases such as muscular dystrophy require early treatment to be most effective, raising questions about coupling diagnosis and treatment. Beyond these examples, many rare genetic disorders will be treatable—in principle—if a streamlined strategy for CRISPR therapeutic development can be implemented¹⁰². With its potential to address unmet medical needs, the clinical use of genome editing will ideally spur changes to regulatory guidelines and cost reimbursement structures that will benefit the field more broadly as these therapies continue to advance.

Notably, all of the genome-editing therapeutics under development aim to treat patients through somatic cell modification. These treatments are designed to affect only the individual who receives the treatment, reflecting the traditional approach to disease mitigation. However, genome editing offers the potential to correct disease-causing mutations in the germline, which would introduce genetic changes that would be passed on to future generations. The scientific and societal challenges associated with human germline editing are distinct from somatic cell editing and are discussed in the next section.

Heritable genome editing

Human germline genome editing can introduce heritable genetic changes in eggs, sperm or embryos. Germline genome editing is already in widespread use in animals and plants, and has been used in human embryos for research purposes. A report of alleged use of human embryo editing that resulted in the birth of twin baby girls with edited genomes has focused global attention on an application of genome editing that must be rigorously regulated, as underscored by international scientific organizations.

Human germline editing differs from somatic cell editing because it results in genetic changes that are heritable if the edited cells are used to initiate a pregnancy (Fig. 4). Germline editing has been used for years in animals, including mice, rats, monkeys and many others,

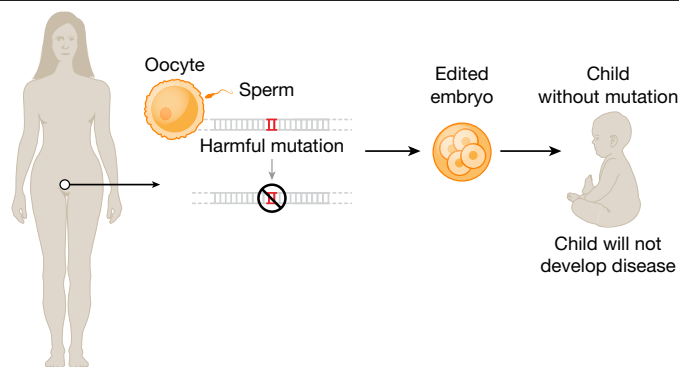


Fig. 4 | Editing the human germline. Genomic changes made during or after embryogenesis may be found in some (mosaic) or all of the cells of the child, including the germline. In contrast to somatic editing (Fig. 1), germline-edited humans can pass these edits down to subsequent generations. In the first human germline-editing experiment in embryos carried to term, the stated goal was to confer HIV resistance, making this example relevant to the real world and highlighting the potential problematic nature of this technique.

and experiments show that it can also be done in both nonviable and viable human embryos^{109–112}. Although none of the published work involves implantation of the edited embryos to initiate a pregnancy, such clinical work was reported at a conference on human genome editing in November 2018, leading to international condemnation in light of clear violations of ethical and scientific guidelines.

This work and the accompanying discussion around human germline editing have raised important questions that affect the future direction of the science as well as the societal and ethical issues that accompany any such applications. First, research using CRISPR–Cas9 in human embryos has challenged our current understanding of DNA repair mechanisms and the developmental pathways that occur in these cells. A report of inaccurate CRISPR–Cas9-based genome editing in non-viable human embryos¹⁰⁹ was not substantiated by later publications, but the mechanism by which double-stranded DNA breaks are repaired in early human embryos remains under debate. Some results were interpreted to indicate repair of a CRISPR–Cas9-targeted gene allele by homology-directed repair with the other allele of the cell as the donor template¹¹³. Other scientists argued that such repair would be impossible given the apparent physical separation of sister chromatids early in embryogenesis, and suggested that the data could also be consistent with large deletions in the embryo genomes^{93,114}. Resolving this fundamental question will require further experiments. Human embryo editing has also begun to reveal differences in the genetics of early development between mice and humans¹¹⁰, underscoring the potential value of research that will be enabled by precision genome modification.

A second question raised by applications of genome editing in human embryos concerns the appropriate professional and societal response. Organizations including the National Academy of Sciences, the National Academy of Medicine, the Royal Society and their equivalents in other countries have sponsored meetings and reports, as have professional societies including the American Society of Human Genetics¹¹⁵, UK Association of Genetic Nurses and Counsellors, Canadian Association of Genetic Counsellors, International Genetic Epidemiology Society, US National Society of Genetic Counselors, American Society for Reproductive Medicine, Asia Pacific Society of Human Genetics, British Society for Genetic Medicine, Human Genetics Society of Australasia, Professional Society of Genetic Counselors in Asia, and Southern African Society for Human Genetics. These groups agree on a number of key points. First, at this time, given the nature and number of unanswered scientific, ethical and policy questions, it is inappropriate to perform germline genome editing that culminates in human pregnancy. Second, in vitro germline genome editing on human

embryos and gametes should be allowed, with appropriate oversight and consent from donors, to facilitate research on the possible future clinical applications of gene editing, and there should be no prohibition on public funding of this research. Third, future clinical applications of human germline genome editing should not proceed unless, at a minimum, there is (a) a compelling medical rationale, (b) an evidence base that supports its clinical use, (c) an ethical justification and (d) a transparent public process to solicit and incorporate stakeholder input.

The third question raised by applications of CRISPR–Cas9 in human embryos is how to move the technology forward while ensuring responsible use. At the time of writing, international commissions convened by the World Health Organization (WHO) and by the US National Academy of Sciences and National Academy of Medicine, together with the Royal Society, are drafting detailed requirements for any potential future clinical use. Medical needs must be defined so that risks versus possible benefits can be evaluated. Most importantly, procedures by which patients could be informed about the technology, its risks and a process for monitoring health outcomes must be determined.

Outlook

Therapeutic genome editing will be realized, at least for some diseases, over the next 5–10 years. This profound opportunity to change healthcare for many people requires scientists, clinicians and bioethicists to work with healthcare economists and regulators to ensure safe, effective and affordable outcomes. The potential impact on patients is too important to wait.

1. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
This study demonstrates dual RNA-programmed DNA cutting by CRISPR–Cas9 and establishes a sgRNA format to direct Cas9 applications, providing a road map for genome editing in human, animal and plant cells.
2. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
4. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
5. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
6. Knott, G. J. & Doudna, J. A. CRISPR–Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
7. Hidalgo-Cantabrana, C., Goh, Y. J. & Barrangou, R. Characterization and repurposing of type I and type II CRISPR–Cas systems in bacteria. *J. Mol. Biol.* **431**, 21–33 (2019).
8. Bao, A. et al. The CRISPR/Cas9 system and its applications in crop genome editing. *Crit. Rev. Biotechnol.* **39**, 321–336 (2019).
9. Terns, M. P. CRISPR-based technologies: impact of RNA-targeting systems. *Mol. Cell* **72**, 404–412 (2018).
10. High, K. A. & Roncarolo, M. G. Gene therapy. *N. Engl. J. Med.* **381**, 455–464 (2019).
11. Pauling, L. et al. Sickle cell anemia, a molecular disease. *Science* **110**, 543–548 (1949).
12. Ingram, V. M. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* **180**, 326–328 (1957).
13. Shieh, P. B. Emerging strategies in the treatment of Duchenne muscular dystrophy. *Neurotherapeutics* **15**, 840–848 (2018).
14. Min, Y.-L., Bassel-Duby, R. & Olson, E. N. CRISPR correction of Duchenne muscular dystrophy. *Annu. Rev. Med.* **70**, 239–255 (2019).
15. Long, C. et al. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345**, 1184–1188 (2014).
16. Miller, J. C. et al. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).
17. Hoban, M. D. et al. Zinc finger nucleases targeting the β -globin locus drive efficient correction of the sickle mutation in CD34⁺ cells. *Blood* **122**, 2904 (2013).
18. Chang, K.-H. et al. Long-term engraftment and fetal globin induction upon BCL11A gene editing in bone-marrow-derived CD34⁺ hematopoietic stem and progenitor cells. *Mol. Ther. Methods Clin. Dev.* **4**, 137–148 (2017).
19. Qasim, W. et al. Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci. Transl. Med.* **9**, eaaj2013 (2017).
20. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
21. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
The authors proposed a surprising but ultimately correct cellular DNA repair mechanism in which double-stranded breaks are enlarged to double-stranded gaps to initiate genetic recombination, forming the basis for genome editing mediated by DNA repair.
22. Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & Jasin, M. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol. Cell. Biol.* **24**, 9305–9316 (2004).
Double-stranded DNA breaks in mammalian cells trigger DNA repair that can introduce site-specific changes in the genome sequence.
23. Liu, J.-J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
24. Yan, W. X. et al. Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell* **70**, 327–339 (2018).
25. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
A DNA-nicking version of CRISPR–Cas9 was fused to a DNA-editing enzyme that enables targeted nucleotide changes to be introduced at Cas9-directed genome locations.
26. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
CRISPR–Cas9 was fused to a DNA-editing enzyme that enables targeted nucleotide editing at genome locations recognized by Cas9, while avoiding double-stranded DNA breaks.
27. Sharon, E. et al. Functional genetic variants revealed by massively parallel precise genome editing. *Cell* **175**, 544–557 (2018).
This study showed that an RNA template can be used together with a Cas9–reverse transcriptase fusion protein to introduce small targeted changes in cellular genomes without involving double-stranded DNA break repair.
28. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
A CRISPR–Cas9–reverse transcriptase fusion protein was used together with extended guide-RNA templates to introduce small sequence changes within approximately 50 base pairs of the location of Cas9 binding.
29. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
This study demonstrated the use of a catalytically deactivated form of CRISPR–Cas9 for transcriptional control in cells.
30. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
31. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
32. Xu, X. & Qi, L. S. A CRISPR–dCas toolbox for genetic engineering and synthetic biology. *J. Mol. Biol.* **431**, 34–47 (2019).
33. Grünwald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
34. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
35. Antoniani, C. et al. Induction of fetal hemoglobin synthesis by CRISPR/Cas9-mediated editing of the human β -globin locus. *Blood* **131**, 1960–1973 (2018).
36. Chung, J. E. et al. CRISPR–Cas9 interrogation of a putative fetal globin repressor in human erythroid cells. *PLoS ONE* **14**, e0208237 (2019).
37. Bjurström, C. F. et al. Reactivating fetal hemoglobin expression in human adult erythroblasts through BCL11A knockdown using targeted endonucleases. *Mol. Ther. Nucleic Acids* **5**, e351 (2016).
38. Liu, N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. *Cell* **173**, 430–442 (2018).
39. Shariati, L. et al. Genetic disruption of the *KLF1* gene to overexpress the γ -globin gene using the CRISPR/Cas9 system. *J. Gene Med.* **18**, 294–301 (2016).
40. Martyn, G. E. et al. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat. Genet.* **50**, 498–503 (2018).
41. Grevet, J. D. et al. Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. *Science* **361**, 285–290 (2018).
42. Martyn, G. E. et al. A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. *Blood* **133**, 852–856 (2019).
43. Lomova, A. et al. Improving gene editing outcomes in human hematopoietic stem and progenitor cells by temporal control of DNA repair. *Stem Cells* **37**, 284–294 (2019).
44. Schumann, K. et al. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Natl Acad. Sci. USA* **112**, 10437–10442 (2015).
45. Dever, D. P. et al. CRISPR/Cas9 β -globin gene targeting in human hematopoietic stem cells. *Nature* **539**, 384–389 (2016).
46. DeWitt, M. A. et al. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* **8**, 360ra134 (2016).
47. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* **25**, 776–783 (2019).
48. Amoasii, L. et al. Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Sci. Transl. Med.* **9**, eaan8081 (2017).
49. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
50. Amoasii, L. et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* **362**, 86–91 (2018).
This article presents evidence that CRISPR–Cas9 can induce corrective genome edits in sufficient cell numbers in vivo to provide therapeutic benefit in a dog model of DMD.
51. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
52. Sharma, R. et al. In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777–1784 (2015).
A potential therapeutic strategy in which blood cells are edited to enable the high-level expression of a desired protein is described.
53. Laoharawee, K. et al. Dose-dependent prevention of metabolic and neurologic disease in murine MPS II by ZFN-mediated in vivo genome editing. *Mol. Ther.* **26**, 1127–1136 (2018).

54. Maeder, M. L. et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* **25**, 229–233 (2019).
This study explored a method for treating inherited retinal disease using a genome-editing approach that uses an AAV5 vector to deliver the S. aureus Cas9 and sgRNAs to photoreceptor cells by subretinal injection.
55. Finn, J. D. et al. A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. *Cell Rep.* **22**, 2227–2235 (2018).
Lipid nanoparticle-based delivery of mRNA-encoded Cas9 and sgRNAs provided therapeutically relevant levels of genome editing in the liver in mice.
56. Hultquist, J. F. et al. A Cas9 ribonucleoprotein platform for functional genetic studies of HIV–host interactions in primary human T cells. *Cell Rep.* **17**, 1438–1452 (2016).
57. Wang, J.-Z., Wu, P., Shi, Z.-M., Xu, Y.-L. & Liu, Z.-J. The AAV-mediated and RNA-guided CRISPR/Cas9 system for gene therapy of DMD and BMD. *Brain Dev.* **39**, 547–556 (2017).
58. Mendell, J. R. et al. Single-dose gene-replacement therapy for spinal muscular atrophy. *N. Engl. J. Med.* **377**, 1713–1722 (2017).
59. Yang, Y. et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).
60. Lau, C.-H. & Suh, Y. In vivo genome editing in animals using AAV–CRISPR system: applications to translational research of human disease. *Flt1000Res.* **6**, 2153 (2017).
61. Bengtsson, N. E. et al. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nat. Commun.* **8**, 14454 (2017).
62. Nelson, C. E. et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403–407 (2016).
In this study, the feasibility of achieving therapeutically meaningful levels of genome editing in affected tissues in a mouse model of muscular dystrophy was demonstrated.
63. Tabebordbar, M. et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407–411 (2016).
The feasibility of achieving therapeutically meaningful levels of genome editing in affected tissues in a mouse model of muscular dystrophy was demonstrated.
64. Li, H. et al. Inhibition of HBV expression in HBV transgenic mice using AAV-delivered CRISPR–SaCas9. *Front. Immunol.* **9**, 2080 (2018).
65. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
66. Charlesworth, C. T. et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).
67. Simhadri, V. L. et al. Prevalence of pre-existing antibodies to CRISPR-associated nuclease Cas9 in the USA population. *Mol. Ther. Methods Clin. Dev.* **10**, 105–112 (2018).
68. Sandoval, I. M., Kuhn, N. M. & Manfredsson, F. P. Multimodal production of adeno-associated virus. *Methods Mol. Biol.* **1937**, 101–124 (2019).
69. Sandro, Q., Relizani, K. & Benchaoui, R. AAV production using baculovirus expression vector system. *Methods Mol. Biol.* **1937**, 91–99 (2019).
70. Strobel, B. et al. Standardized, scalable, and timely flexible adeno-associated virus vector production using frozen high-density HEK-293 cell stocks and CELLdiscs. *Hum. Gene Ther. Methods* **30**, 23–33 (2019).
71. Miller, J. B. et al. Non-viral CRISPR/Cas gene editing in vitro and in vivo enabled by synthetic nanoparticle co-delivery of Cas9 mRNA and sgRNA. *Angew. Chem. Int. Edn* **56**, 1059–1063 (2017).
72. Wang, M. et al. Efficient delivery of genome-editing proteins using bio-reducible lipid nanoparticles. *Proc. Natl Acad. Sci. USA* **113**, 2868–2873 (2016).
73. Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* **9**, 2184 (2018).
74. Gao, X. et al. Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).
75. Zuris, J. A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).
76. Staahl, B. T. et al. Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).
77. Ding, Y. et al. Gold nanoparticles for nucleic acid delivery. *Mol. Ther.* **22**, 1075–1083 (2014).
78. Glass, Z., Li, Y. & Xu, Q. Nanoparticles for CRISPR–Cas9 delivery. *Nat. Biomed. Eng.* **1**, 854–855 (2017).
79. Lee, K. et al. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* **1**, 889–901 (2017).
80. Sago, C. D. et al. High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. *Proc. Natl Acad. Sci. USA* **115**, E9944–E9952 (2018).
81. Qin, W. & Wang, H. Delivery of CRISPR–Cas9 into mouse zygotes by electroporation. *Methods Mol. Biol.* **1874**, 179–190 (2019).
82. Tanihara, F. et al. Generation of a TP53-modified porcine cancer model by CRISPR/Cas9-mediated gene modification in porcine zygotes via electroporation. *PLoS ONE* **13**, e0206360 (2018).
83. Xu, L. et al. CRISPR-mediated genome editing restores dystrophin expression and function in mdx mice. *Mol. Ther.* **24**, 564–569 (2016).
84. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
This study demonstrated the use of purified protein–guide RNA complexes for genome editing in human cells.
85. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
86. Gaj, T. et al. Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery. *Nucleic Acids Res.* **45**, e98 (2017).
87. Rouet, R. et al. Receptor-mediated delivery of CRISPR–Cas9 endonuclease for cell-type-specific gene editing. *J. Am. Chem. Soc.* **140**, 6596–6603 (2018).
88. Yin, J. et al. Potent protein delivery into mammalian cells via a supercharged polypeptide. *J. Am. Chem. Soc.* **140**, 17234–17240 (2018).
89. Riley, R. S., June, C. H., Langer, R. & Mitchell, M. J. Delivery technologies for cancer immunotherapy. *Nat. Rev. Drug Discov.* **18**, 175–196 (2019).
90. Sun, Y. et al. Advances of blood cell-based drug delivery systems. *Eur. J. Pharm. Sci.* **96**, 115–128 (2017).
91. Sharma, P. et al. Efficient intracellular delivery of biomacromolecules employing clusters of zinc oxide nanowires. *Nanoscale* **9**, 15371–15378 (2017).
92. Kim, D., Luk, K., Wolfe, S. A. & Kim, J.-S. Evaluating and enhancing target specificity of gene-editing nucleases and deaminases. *Annu. Rev. Biochem.* **88**, 191–220 (2019).
93. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
94. Poirot, L. et al. Multiplex genome-edited T-cell manufacturing platform for “off-the-shelf” adoptive T-cell immunotherapies. *Cancer Res.* **75**, 3853–3864 (2015).
95. Bak, R. O. et al. Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *eLife* **6**, e27873 (2017).
96. Tichy, E. D. et al. Mouse embryonic stem cells, but not somatic cells, predominantly use homologous recombination to repair double-strand DNA breaks. *Stem Cells Dev.* **19**, 1699–1711 (2010).
97. Buechele, C. et al. MLL leukemia induction by genome editing of human CD34⁺ hematopoietic cells. *Blood* **126**, 1683–1694 (2015).
98. Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423–427 (2014).
99. NIST. *NIST Genome Editing Consortium* <https://www.nist.gov/programs-projects/nist-genome-editing-consortium> (NIST, 2017).
100. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
101. Liu, X. S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247 (2016).
CRISPR–Cas9 fusion proteins were used to introduce targeted epigenetic changes into cellular genomes.
102. Wilson, R. C. & Carroll, D. The daunting economics of therapeutic genome editing. *CRISPR J.* **2**, 280–284 (2019).
103. Wagner, D. L. et al. High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* **25**, 242–248 (2019).
104. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
105. Ihry, R. J. et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
106. Schirolli, G. et al. Precise gene editing preserves hematopoietic stem cell function following transient p53-mediated DNA damage response. *Cell Stem Cell* **24**, 551–565.e8 (2019).
107. Xu, L. et al. CRISPR-edited stem cells in a patient with HIV and acute lymphocytic leukemia. *N. Engl. J. Med.* **381**, 1240–1247 (2019).
108. Porteus, M. H. A new class of medicines through DNA editing. *N. Engl. J. Med.* **380**, 947–959 (2019).
109. Liang, P. et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell* **6**, 363–372 (2015).
110. Fogarty, N. M. E. et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67–73 (2017).
111. Tang, L. et al. CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein. *Mol. Genet. Genomics* **292**, 525–533 (2017).
112. Ma, H. et al. Correction of a pathogenic gene mutation in human embryos. *Nature* **548**, 413–419 (2017).
113. Kaul, S., Heitner, S. B. & Mitalipov, S. Sarcomere gene mutation correction. *Eur. Heart J.* **39**, 1506–1507 (2018).
114. Egli, D. et al. Inter-homologue repair in fertilized human eggs? *Nature* **560**, E5–E7 (2018).
115. Ormond, K. E. et al. Human germline genome editing. *Am. J. Hum. Genet.* **101**, 167–176 (2017).

Acknowledgements J.A.D. thanks M. Hochstrasser and M. Triplet for extensive expert assistance with manuscript editing, formatting, referencing and illustrations. D. Carroll, F. Urnov and R. Wilson provided comments on the manuscript. T. Tolpa created the artwork with input from M. Hochstrasser and E. Stahl, and with support from the Innovative Genomics Institute. J.A.D. is an investigator of the Howard Hughes Medical Institute (HHMI) and a Paul Allen Distinguished Investigator. In addition to funding from HHMI and the Paul Allen Frontiers Group, research in the Doudna laboratory is supported by the Defense Advanced Research Projects Agency (DARPA) (award HRO011-17-2-0043), the William M. Keck Foundation, a Collaborative MS Research Center Award from the National Multiple Sclerosis Society, the Centers for Excellence in Genomic Science of the National Institutes of Health under award number RM1HG009490, the Somatic Cell Genome Editing Program of the Common Fund of the National Institutes of Health under award number U01AI142817-02 and the National Science Foundation under award number 1817593.

Competing interests J.A.D. is a co-founder of Caribou Biosciences, Editas Medicine, Intellia Therapeutics, Scribe Therapeutics and Mammoth Biosciences; a scientific adviser to Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Synthego, Inari and eFFECTOR Therapeutics; and a director of Johnson & Johnson. The Regents of the University of California have patents issued and pending for CRISPR-related technologies on which J.A.D. is an inventor.

Additional information

Correspondence and requests for materials should be addressed to J.A.D.

Peer review information *Nature* thanks R. Alta Charo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Galactic-scale gas wave in the solar neighbourhood

<https://doi.org/10.1038/s41586-019-1874-z>

Received: 20 June 2019

Accepted: 24 October 2019

Published online: 7 January 2020

João Alves^{1,2*}, Catherine Zucker³, Alyssa A. Goodman^{2,3}, Joshua S. Speagle³, Stefan Meingast¹, Thomas Robitaille⁴, Douglas P. Finkbeiner^{3,5}, Edward F. Schlafly⁶ & Gregory M. Green⁷

For the past 150 years, the prevailing view of the local interstellar medium has been based on a peculiarity known as the Gould Belt^{1–4}, an expanding ring of young stars, gas and dust, tilted about 20 degrees to the Galactic plane. However, the physical relationship between local gas clouds has remained unknown because the accuracy in distance measurements to such clouds is of the same order as, or larger than, their sizes^{5–7}. With the advent of large photometric surveys⁸ and the astrometric survey⁹, this situation has changed¹⁰. Here we reveal the three-dimensional structure of all local cloud complexes. We find a narrow and coherent 2.7-kiloparsec arrangement of dense gas in the solar neighbourhood that contains many of the clouds thought to be associated with the Gould Belt. This finding is inconsistent with the notion that these clouds are part of a ring, bringing the Gould Belt model into question. The structure comprises the majority of nearby star-forming regions, has an aspect ratio of about 1:20 and contains about three million solar masses of gas. Remarkably, this structure appears to be undulating, and its three-dimensional shape is well described by a damped sinusoidal wave on the plane of the Milky Way with an average period of about 2 kiloparsecs and a maximum amplitude of about 160 parsecs.

To reveal the physical connections between clouds in the local interstellar medium (ISM), we determined the three-dimensional (3D) distribution of all local cloud complexes¹¹ by deriving accurate distances to about 380 lines of sight. The lines of sight were chosen to include not only all known local clouds^{10,12} but also potential bridges between them, as traced by lower-column-density gas. Figure 1 presents the distribution of lines of sight studied towards the Galactic anti-centre and illustrates our overall approach. Each line of sight covers an area in the sky of about 450 arcmin² and includes both foreground and background stars for a particular direction towards a cloud. The distances and the colours of these stars are used to compute a distance to the cloud (see Methods).

In the interactive figure in Supplementary Information we present the distribution of cloud distances to all of the studied lines of sight in a Cartesian XYZ frame where *X* increases towards the Galactic centre, *Y* increases along the direction of rotation of the Galaxy and *Z* increases upwards out of the Galactic plane. In the *X–Y* projection (a top-down view of the Galactic disk), it is clear that cloud complexes are not randomly distributed, but instead tend to form elongated and relatively linear arrangements. Surprisingly, we find that one of the nearest structures, at about 300 pc from the Sun at its closest point, is exceptionally straight and narrow in the *X–Y* plane. This straight structure: (1) undulates systematically in the *Z* axis for about 2.7 kpc on the *X–Y* plane, (2) is co-planar in essentially its entire extent and (3) displays radial velocities¹³ indicating that the structure is not a random

alignment of molecular cloud complexes but a kinematically coherent structure. We find that this structure is well modelled as a damped sinusoidal wave. The red points in Fig. 2 were selected by the fitting procedure, by explicitly modelling inliers and outliers. We tested the validity of the model by modelling the ‘tenuous connections’ separately and confirming that they meet the same inlier criteria that were first applied to the major clouds. For more details on the statistical modelling, see Methods.

Apart from the continuous undulating 3D distribution, there is also very limited kinematic evidence that the structure is physically oscillating around the mid-plane of the Galaxy, as any sinusoidal mass distribution centred on the Galactic plane should. The Galactic space velocities (*U*, *V*, *W*) in the local-standard-of-rest frame for a sample of young stellar objects associated with the Orion A cloud near the ‘trough’ of this structure are (–10.2, –1.2, –0.1) km s^{–1} (J. Grossschedl, private communication), implying that Orion A has now reached its maximum distance from the Galactic plane before falling back into the plane. These observations also indicate that Orion, and probably the large structure described here, is moving tangentially with about the same speed as the local Galactic disk.

This spatially and kinematically coherent structure has an amplitude of roughly 160 pc at its maximum and a period of roughly 2 kpc. We estimate the mass of the structure to be at least $3 \times 10^6 M_{\odot}$ (M_{\odot} , solar mass) by integrating the Planck opacity map¹⁴ for the different cloud complexes in the structure at their estimated distances. The procedures

¹University of Vienna, Department of Astrophysics, Vienna, Austria. ²Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA, USA. ³Department of Astronomy, Harvard University and Center for Astrophysics, Harvard and Smithsonian, Cambridge, MA, USA. ⁴Aperio Software, Leeds, UK. ⁵Department of Physics, Harvard University, Cambridge, MA, USA.

⁶Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁷Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA, USA. *e-mail: joao.alves@univie.ac.at

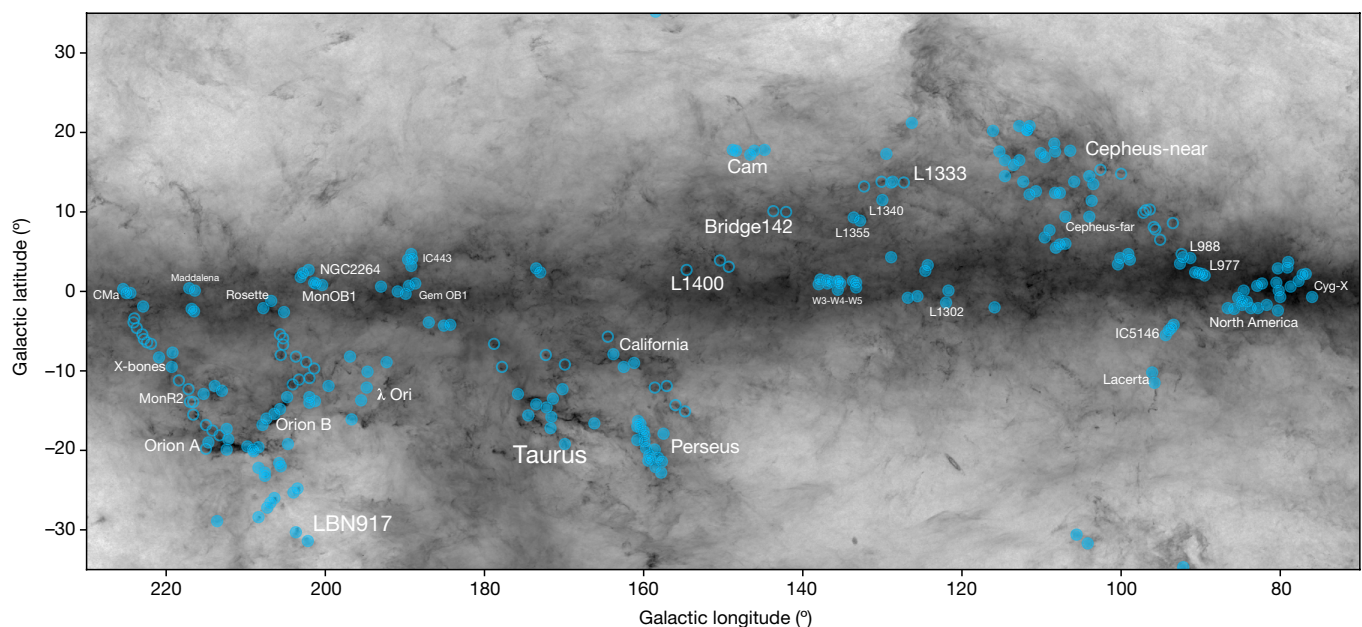


Fig. 1 | Sky map of targeted star-forming regions towards the anti-centre of the Milky Way. The filled circles represent the studied lines of sight used to determine accurate distances to known nearby star-forming complexes (the sizes of the region labels are roughly proportional to their distance). The

open circles represent lines of sight towards lower-column-density envelopes between complexes. The background greyscale map shows the column density distribution derived from Planck data¹⁴.

used to compute the mass and model the 3D shape of the structure are described in Methods. We name the structure the Radcliffe Wave in honour of both the early-20th-century female astronomers from Radcliffe College and the interdisciplinary spirit of the current Radcliffe Institute, which contributed to this discovery. The structure can also be seen at lower resolution in recent all-sky 3D dust maps^{15–17} (see Fig. 2). A second linear structure, the ‘split’¹⁶, is about 1 kpc long and seems to contain the Sco-Cen, Aquila and Serpens clouds, as well as a previously

unidentified complex. The functional form of the split is different, however, in that it is largely confined to the Galactic plane over much of its length and does not seem to be undulating.

The interactive figure in Supplementary Information, which displays the 3D location of the Gould Belt¹⁸, illustrates that with the improved distances, this structure is a poor fit to the data, which comprise only clouds from Sco-Cen and Orion—the traditional anchors of the Gould Belt. This fact alone challenges the existence of a belt, as two points can

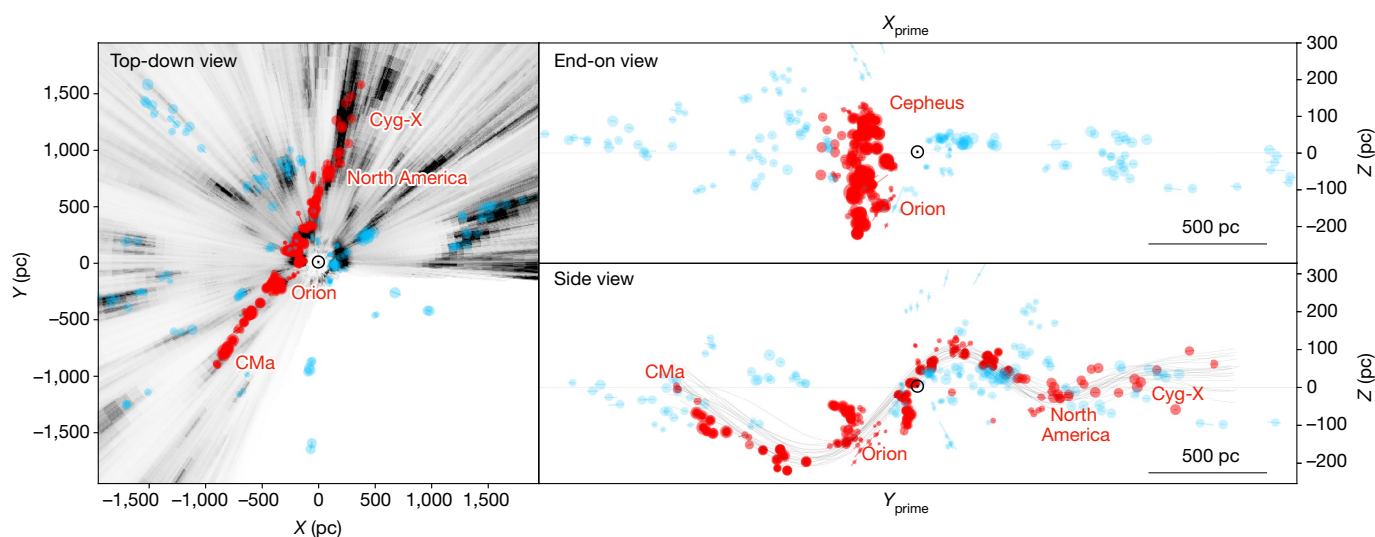


Fig. 2 | 3D distribution of local clouds. The position of the Sun is marked with \odot . The size of the symbols is proportional to the column density. The red points were selected by a fitting algorithm, as described in Methods. These describe a spatially and kinematically coherent structure that we term the Radcliffe Wave (possible models are shown by the grey lines in the bottom-right panel). The greyscale map in the left panel shows an integrated dust map¹⁷ ($-300 \text{ pc} < Z < 300 \text{ pc}$), which indicates that our sample of cloud distances is essentially complete. To highlight the undulation and co-planarity of the structure, the right panels show projections in which the X - Y frame has been rotated anticlockwise by 33°

(top, $X_{\text{prime}}-Z$) and clockwise by 120° (bottom, $Y_{\text{prime}}-Z$) for an observer facing the Galactic anti-centre. The 1σ statistical uncertainties on the distance (usually 1–2%) are represented by line segments that are usually smaller than the symbols. There is an additional systematic uncertainty on the distance, which is estimated¹⁰ to be about 5%. For an interactive version of this figure, including additional layers not shown here (for example, a model of the Gould Belt and log-spiral arm fits), see Supplementary Information and https://faun.rc.fas.harvard.edu/czucker/Paper_Figures/radwave.html.

always define a ring. Because four out of five of the Gould Belt clouds (Orion, Perseus, Taurus, Cepheus) are part of the much larger Radcliffe Wave, whereas one of the five (Ophiuchus) is part of the split, we propose that the Gould Belt is a projection effect of two linear cloud complexes against the sky. Our results provide an alternate explanation for the 20° inclination of the Gould Belt: it is simply the orientation of the Radcliffe Wave from trough (Orion) to crest (Cepheus). With these considerations, the X – Y distribution of local B-stars in these regions from the 30-year-old Hipparcos satellite¹⁹ resembles the two elongated linear structures in Fig. 2 more closely than a ring, bolstering previous suspicions that the Gould Belt is a projection effect²⁰.

In Supplementary Information, one can access an interactive display of the 3D location of the Local Arm of the Milky Way as traced by masers²¹ and investigate the relation between the Radcliffe Wave and the Local Arm. The Radcliffe Wave (red points) is about 20% of the width and 40% of the length of the Local Arm²² and makes up for an important fraction of the Local Arm's mass and number of cloud complexes. On the other hand, the Local Arm is much more dispersed and includes local complexes that are not part of the Radcliffe Wave (for example, Mon OB1, California, Cepheus Far and Ophiuchus). Whereas there is excellent agreement between our distance measurements and maser-defined distances¹², the log-spiral fit of the maser data crosses the Radcliffe Wave at an angle of about 25°. The mismatch between the Radcliffe Wave and the log-spiral fit suggests that the Local Arm is more structured and complex than previously thought, but is consistent with arms being composed of quasi-linear structures on kiloparsec scales^{23,24}.

The origin of the Radcliffe Wave is unclear. The structure is too large (and too straight) to have formed by the feedback of a previous generation of massive stars. More probably, this narrow structure is the outcome of a large-scale Galactic process of gas accumulation, either from a shock front in a spiral arm²⁵ or from gravitational settling and cooling on the plane of the Milky Way (Kim, W.-T. & Ostriker, E. C., manuscript in preparation). Linear kiloparsec-sized structures similar to the one presented here have been seen in nearby galaxies²⁶ and in numerical simulations²⁴ of spiral-arm formation.

The undulation of the Radcliffe Wave is even harder to explain. The accretion of a tidally stretched gas cloud settling into the Galactic disk could in principle mimic the shape and the damped undulation of the structure, but it requires synchronization with the Galactic rotation (Orion's velocity in the local-standard-of-rest frame, $V_{\text{LSR}} \approx 0 \text{ km s}^{-1}$), which is plausible but seems unlikely. Analogous kiloparsec-sized waves (or corrugations) have been seen in nearby galaxies²⁷ with amplitudes similar to the undulation seen in Fig. 2 (ref. ²⁸), but their origins often call for perturbers. Identifying possible disruption events, their corresponding progenitors and their relationship to the Radcliffe Wave is a substantial challenge that should be explored.

Our findings call for a revision of the architecture of gas in the solar neighbourhood and a re-interpretation of phenomena that are generally associated with the Gould Belt, such as the Lindblad ring and the Cas–Tau/ α -Per populations, among many others²⁹. The Radcliffe Wave provides a framework for understanding molecular cloud formation and evolution. Follow-up work, in particular on the kinematics of this structure, will provide insights into the relative roles of gravity, feedback and magnetic fields in star-formation research.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1874-z>.

- Herschel, J. F. W. *Results of Astronomical Observations Made During the Years 1834, 5, 6, 7, 8, at the Cape of Good Hope; Being the Completion of a Telescopic Survey of the Whole Surface of the Visible Heavens, Commenced in 1825* (Smith, Elder and Company, 1847).
- Gould, B. A. On the number and distribution of the bright fixed stars. *Am. J. Sci.* **38**, 325–333 (1874).
- Bobylev, V. V. The Gould belt. *Astrophysics* **57**, 583–604 (2014).
- Palouš, J. & Ehlerová, S. in *Handbook of Supernovae* (eds Alsabti, A. W. & Murdin, P.) 2301–2311 (Springer, 2016).
- Maddalena, R. J., Morris, M., Moscovitz, J. & Thaddeus, P. The large system of molecular clouds in Orion and Monoceros. *Astrophys. J.* **303**, 375–391 (1986).
- Lombardi, M., Lada, C. J. & Alves, J. Hipparcos distance estimates of the Ophiuchus and the Lupus cloud complexes. *Astron. Astrophys.* **480**, 785–792 (2008).
- Schlaflly, E. F. et al. A large catalog of accurate distances to molecular clouds from ps1 photometry. *Astrophys. J.* **786**, 29 (2014).
- Chambers, K. C. et al. *The Pan-STARRS1 surveys*. Preprint at <https://arxiv.org/abs/1612.05560> (2016).
- Brown, A. G. A., Vallenari, A., Prusti, T. & de Bruijn, J. H. J. Gaia Data Release 2: summary of the contents and survey properties. *Astron. Astrophys. Suppl. Ser.* **616**, A1 (2018).
- Zucker, C. et al. A large catalog of accurate distances to local molecular clouds: the Gaia DR2 edition. *Astrophys. J.* **879**, 125 (2019).
- Reipurth, B. (ed.) *Handbook of Star Forming Regions, Volume I: The Northern Sky* Vol. 4 (ASP, 2008).
- Zucker, C. et al. A compendium of distances to molecular clouds in the star formation handbook. *Astron. Astrophys.* **633**, A51 (2020).
- Dame, T. M., Hartmann, D. & Thaddeus, P. The Milky Way in molecular clouds: a new complete CO survey. *Astrophys. J.* **547**, 792–813 (2001).
- Planck Collaboration Planck 2013 results. XI. All-sky model of thermal dust emission. *Astron. Astrophys.* **571**, A11 (2014).
- Green, G. M. et al. Galactic reddening in 3D from stellar photometry – an improved map. *Mon. Not. R. Astron. Soc.* **478**, 651–666 (2018).
- Lallement, R. et al. Gaia-2MASS 3D maps of Galactic interstellar dust within 3 kpc. *Astron. Astrophys.* **625**, A135 (2019).
- Green, G. M., Schlaflly, E. F., Zucker, C., Speagle, J. S. & Finkbeiner, D. P. A 3D dust map based on gaia, Pan-STARRS 1 and 2MASS. *Astrophys. J.* **887**, 93 (2019).
- Perrot, C. A. & Grenier, I. A. 3D dynamical evolution of the interstellar gas in the Gould belt. *Astron. Astrophys. Suppl. Ser.* **404**, 519–531 (2003).
- Elias, F., Cabrera-Caño, J. & Alfaro, E. J. OB stars in the solar neighborhood. I. Analysis of their spatial distribution. *Astron. J.* **131**, 2700–2709 (2006).
- Bouy, H. & Alves, J. F. Cosmography of OB stars in the solar neighbourhood. *Astron. Astrophys. Suppl. Ser.* **584**, A26 (2015).
- Reid, M. J., Dame, T. M., Menten, K. M. & Brunthaler, A. A parallax-based distance estimator for spiral arm sources. *Astrophys. J.* **823**, 77 (2016).
- Reid, M. J. et al. Trigonometric parallaxes of high mass star forming regions: the structure and kinematics of the Milky Way. *Astrophys. J.* **783**, 130 (2014).
- Honig, Z. N. & Reid, M. J. Characteristics of spiral arms in late-type galaxies. *Astrophys. J.* **800**, 53 (2015).
- D'Onghia, E., Vogelsberger, M. & Hernquist, L. Self-perpetuating spiral arms in disk galaxies. *Astrophys. J.* **766**, 34 (2013).
- Goodman, A. A. et al. The bones of the Milky Way. *Astrophys. J.* **797**, 53 (2014).
- Elmegreen, B. G., Elmegreen, D. M. & Efremov, Y. N. Regularly spaced infrared peaks in the dusty spirals of Messier 100. *Astrophys. J.* **863**, 59 (2018).
- Edelsohn, D. J. & Elmegreen, B. G. Corrugations in galactic discs generated by magellanic-type perturbers. *Mon. Not. R. Astron. Soc.* **287**, 947–954 (1997).
- Matthews, L. D. & Uson, J. M. Corrugations in the disk of the edge-on spiral galaxy IC 2233. *Astrophys. J.* **688**, 237–244 (2008).
- Bally, J. in *Handbook of Star Forming Regions* Vol. 4 (ed. Reipurth, B.) 459–370 (Astronomical Society of the Pacific, 2008).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Distances

Distances were determined for 326 lines of sight using major local molecular clouds and 54 ‘bridging’ lines of sight in between molecular clouds coincident with the projected structure of the Radcliffe Wave. The methodology used to obtain the distances and the full catalogue of lines of sight for the major clouds are presented in complementary work^{10,12}. Lines of sight for the major clouds were chosen to coincide with star-forming regions presented in ref. ¹¹, which is considered to be the most comprehensive resource on individual low- and high-mass star-forming regions out to 2 kpc. Lines of sight for the tenuous connections were chosen in two dimensions to coincide with structures (for example, diffuse filamentary ‘bridges’; see Fig. 1) that appeared to span the known star-forming regions on the plane of the sky without a priori knowledge of their distances. These were later used to validate the 3D modelling, which did not incorporate these distances.

Mass

We estimate the mass of the Radcliffe Wave to be about $3 \times 10^6 M_\odot$ using the Planck column density map shown in Fig. 1. To estimate the total mass, we first define the extent and depth for each complex in Fig. 1 using the information on the line-of-sight distances. We then integrate the column density map using the average distance to each complex. To correct for background contamination, which is critical for complexes closer to the plane, we subtract an average column density per complex estimated at the same Galactic latitude. Our resulting mass estimate of the Radcliffe Wave is probably an approximate lower limit to the true mass of the structure, given that the regions of the wave crossing the plane from Perseus to Cepheus and from Cepheus to Cygnus are poorly sampled owing to Galactic plane confusion.

Kinematics

We apply the open-source Gaussian fitting package PySpecKit³⁰ over local ¹²CO spectral observations¹³ to obtain the observed velocities of the star-forming regions shown in Extended Data Fig. 1. For each line of sight, we compute a spectrum over the same region that is used to compute the dust-based distances. We then fit a single-component Gaussian to each spectrum and assign the mean value as the velocity. We are not able to derive observed velocities for ~25% of the sample that either fall outside the boundaries of the survey¹³, have no appreciable emission above the noise threshold and/or contain spectra that are not well modelled by a single-component Gaussian. The spectra that are not well modelled by a single-component Gaussian represent about 2% of the lines of sight and occur towards the most massive, structured and extinguished lines of sight in the sample, suggesting that these spectra could contain CO self-absorption features. We have confirmed that these more complex spectra do not show evidence of multiple distance components. Regardless, because the predicted velocities rely only on the estimated cloud distances assuming that they follow the ‘universal’ Galactic rotation curve²², not every line of sight in Extended Data Fig. 1 has a corresponding observed velocity associated with its predicted velocity.

We compute the background grayscale map in Extended Data Fig. 1 by collapsing the ¹²CO spectral observations over only the regions that are coincident with the cloud lines of sight on the plane of the sky.

3D modelling

We model the centre of the Radcliffe Wave using a quadratic function with respect to X , Y and Z specified by three sets of ‘anchor points’, (x_0, y_0, z_0) , (x_1, y_1, z_1) and (x_2, y_2, z_2) . We find that a simpler linear function is unable to accurately model the observed curvature in the structure and is subsequently disfavoured by the data.

The undulating behaviour with respect to the centre is described by a damped sinusoidal function relative to the X - Y plane with a decaying period and amplitude, which we parameterize as

$$\Delta z(t) = A \times \exp\left[-\delta\left(\frac{d(t)}{1 \text{ kpc}}\right)^2\right] \times \sin\left[\left(\frac{2\pi d(t)}{P}\right)\left(1 + \frac{d(t)/d_{\max}}{\gamma}\right) + \phi\right] \quad (1)$$

where $d(t) = \|(x, y, z)(t) - (x_0, y_0, z_0)\| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$ is the Euclidean distance from the start of the wave parameterized by t , d_{\max} is the distance at the end of the wave, A is the amplitude, P is the period, ϕ is the phase, δ sets the rate at which the amplitude decays and γ sets the rate at which the period decays. We explored introducing an additional parameter to account for rotation around the primary axis determined by our quadratic fit, but found that the results were entirely consistent with the structure oscillating in the X - Y plane, and so excluded this parameter in our final model.

We assume the distance of each cloud d_{cloud} relative to our model to be normally distributed with an unknown scatter σ that is roughly equivalent to the radius of the wave. To account for different positions along the wave, we define this distance relative to the closet point as

$$d_{\text{cloud}} = \min_t \left(\|(x_{\text{cloud}}, y_{\text{cloud}}, z_{\text{cloud}}) - (x_{\text{wave}}, y_{\text{wave}}, z_{\text{wave}})(t)\| \right) \quad (2)$$

Finally, we account for structure ‘off’ the wave by fitting a mixture model. We assume that a fraction f of clouds are distributed quasi-uniformly in a volume of roughly 10^7 pc^3 , so that the remaining $1 - f$ is part of the wave. We treat f entirely as a nuisance parameter because it is completely degenerate with the volume of our uniform outlier model, although we have specified it so that the uniform component will contribute a ‘minority’ of the fit (<40%).

Assuming that the distances to each of our n clouds have been derived independently, and defining $\theta = \{x_0, y_0, z_0, x_1, y_1, z_1, x_2, y_2, z_2, P, A, \phi, \gamma, \delta, \sigma, f\}$, the likelihood for a given realization of our 16-parameter 3D model is

$$\mathcal{L}(\theta) = \prod_{i=1}^n \left[(1-f) \mathcal{L}_{\text{cloud},i}(\theta) + f \mathcal{L}_{\text{unif},i}(\theta) \right] \quad (3)$$

where

$$\mathcal{L}_{\text{cloud},i}(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{d_{\text{cloud},i}^2}{\sigma^2}\right] \quad (4)$$

$$\mathcal{L}_{\text{unif},i}(\theta) = 10^{-7}$$

We infer the posterior probability distribution $\mathcal{P}(\theta)$ of the 3D model parameters to be consistent with our cloud distances (excluding all bridging features) using Bayes’ theorem:

$$\mathcal{P}(\theta) \propto \mathcal{L}(\theta)\pi(\theta) \quad (5)$$

where $\pi(\theta)$ is our prior distributions over the parameters of interest. We set our prior $\pi(\theta)$ to be independent for each parameter, on the basis of initial fits. The priors on each parameter are described in Extended Data Table 1, where $\mathcal{N}(\mu, \sigma)$ is a normal distribution with mean μ and standard deviation σ and $\mathcal{U}(a, b)$ is a uniform distribution with lower bound a and upper bound b .

We generate samples from $\mathcal{P}(\theta)$ with the nested sampling code dynesty³¹ using a combination of uniform sampling with multi-ellipsoid decompositions and 1,000 live points. A summary of the derived properties of the Radcliffe Wave are listed in Extended Data Table 2 along with their associated 95% credible intervals. The 20 random samples from $\mathcal{P}(\theta)$ are plotted in Fig. 2 to illustrate the uncertainties in our model.

Using our samples, we associate particular sightlines with the Wave by computing the mean odds ratio averaged over our posterior

$$\langle R_i \rangle = \int \frac{(1-f)\mathcal{L}_{\text{cloud},i}(\theta)}{f\mathcal{L}_{\text{unif},i}(\theta)} \mathcal{P}(\theta) d\theta \quad (6)$$

based on our set of samples. We subsequently classify all objects with $\langle R_i \rangle > 1$ as being part of the Radcliffe Wave, which is used as the criterion for associating sources in Fig. 2. We find that this condition holds true for 43% of the sources used to determine our initial model. Our overall conclusions do not change if larger, more selective thresholds are chosen.

As further validation, we subsequently compute $\langle R_i \rangle$ for each of the 54 bridging lines of sight targeted to follow the projected structure of the Radcliffe Wave. We find that all 54 lines of sight satisfy our $\langle R_i \rangle > 1$ condition, further confirming the continuous nature of the Wave between individual clouds.

In addition to the parameters derived above, we estimate the total length of the feature in our dataset by computing the line integral along our model from the clouds at the endpoints, finding a length of 2.7 ± 0.2 kpc (95% credible interval). The derived physical properties of the feature are listed in Extended Data Table 3.

Data availability

The datasets generated and/or analysed during the current study are publicly available on the Harvard Dataverse: the distances to the major star-forming clouds are available at <https://doi.org/10.7910/DVN/07L7YZ> and the tenuous connections at <https://doi.org/10.7910/DVN/K16GQX>.

Code availability

The software used to determine the distances to star-forming regions is publicly available on Zenodo (<https://doi.org/10.5281/zenodo.3348370> and <https://doi.org/10.5281/zenodo.3348368>). The code used for model fitting is available from J.S.S. (jspeagle@cfa.harvard.edu) on reasonable request.

30. Ginsburg, A. PySpecKit: Python spectroscopic toolkit. Astrophysics Source Code Library ascl:1109.001 (2011).
31. Speagle, J.S. *dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences*. Preprint at <https://arxiv.org/abs/1904.02180> (2019).

Acknowledgements J.A. thanks the Radcliffe Institute, where this work was developed, and where J.A. discovered the work of visual artist A. von Merten on H. Leavitt's work, which inspired us to "see more". We acknowledge the organizers and participants of the 'The Milky Way in the age of Gaia' workshop of the 2018 Paris-Saclay International Programs for Physical Sciences, as well as the Interstellar Institute, for discussions at the early stage of this work. We benefited from discussions with T. Dame, M. Reid, A. Burkert and M. Davies. J.A. acknowledges the TURIS and Data Science Research Platforms of the University of Vienna. C.Z. and J.S.S. are supported by the NSF Graduate Research Fellowship Program (grant number 1650114) and the Harvard Data Science Initiative. D.P.F. and C.Z. acknowledge support by NSF grant AST-1614941. E.F.S. acknowledges support by NASA through ADAP grant NNN17AE751 and Hubble Fellowship grant HST-HF2-51367.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. The computations in this paper used resources from the Odyssey cluster, which is supported by the FAS Division of Science Research Computing Group at Harvard University. The high-dimensional visualization software Glue, which was used to explore, visualize and understand the Radcliffe Wave, was created by A.A.G., T.R., C.Z. and others, and has been supported by US Government contract NAS5-03127 through NASA's James Webb Space Telescope Mission and NSF awards OAC-1739657 and AST-1908419. We are grateful to A. Johnson and others at Plotly Graphing Library for their help creating the 3D interactive figure, which was output from Glue to Plotly. WorldWide Telescope (WWT), which was used within Glue to visualize the wave, is currently supported by NSF grant 1642446 to the American Astronomical Society. WWT was originally created by C. Wong and J. Fay at Microsoft Research, which supported WWT development before the American Astronomical Society. J.S.S. thanks R. Bleich, and J.A. thanks A. dell'Erba, J. Alves, M. Alves and R. Alves for continuing support.

Author contributions J.A. led the work and wrote most of the text. All authors contributed to the writing of the manuscript. C.Z. and J.S.S. led the data analysis and distance modelling with E.F.S., G.M.G. and D.P.F. C.Z. and J.A. led the kinematics analysis. J.A., C.Z. and A.A.G. led the visualization efforts. J.S.S. led the 3D modelling. J.A., C.Z. and A.A.G. led the efforts to interpret the results. T.R., A.A.G., J.S.S. and C.Z. contributed to the development of the software used in this work.

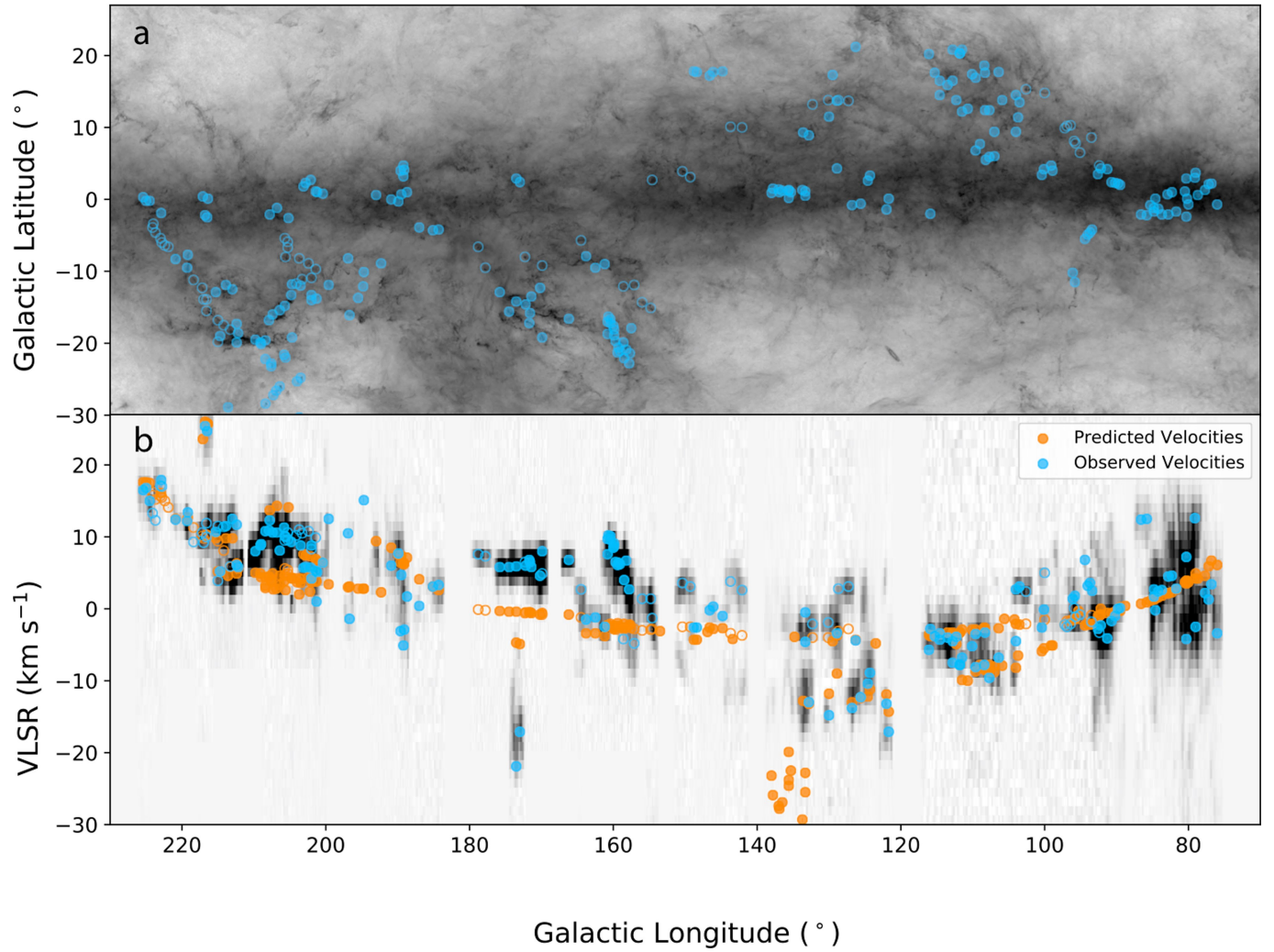
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1874-z>.

Correspondence and requests for materials should be addressed to J.A.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Position-velocity diagram. a, b, The blue points in **a** are as in Fig. 1 and the orange points in **b** represent the predicted positions of the blue points as if they were following a 'universal' Galactic rotation curve²². The line segments represent 1σ errors, derived from a Gaussian fitting for the observed velocities and the distance uncertainties for the predicted velocities, and are generally smaller than the symbols. The quasi-linear arrangement in

velocity of the Radcliffe Wave complexes suggests that the structure is not a random alignment of molecular cloud complexes, but a kinematically coherent structure. The tentative decoupling between observed and predicted velocities also indicate that the Radcliffe Wave is a kinematically coherent structure. VLSR, velocity in the local-standard-of-rest frame.

Extended Data Table 1 | Priors on Radcliffe Wave parameters

Parameter	Prior	Parameter	Prior
x_0	$\mathcal{N}(-900 \text{ pc}, 100 \text{ pc})$	P	$\mathcal{N}(3500 \text{ pc}, 300 \text{ pc})$
y_0	$\mathcal{N}(-900 \text{ pc}, 100 \text{ pc})$	A	$\mathcal{N}(170 \text{ pc}, 20 \text{ pc})$
z_0	$\mathcal{N}(0 \text{ pc}, 50 \text{ pc})$	ϕ	$\mathcal{N}(2.9 \text{ rad}, 0.5 \text{ rad})$
x_1	$\mathcal{N}(-300 \text{ pc}, 100 \text{ pc})$	$\ln \gamma$	$\mathcal{N}(-0.5, 0.5)$
y_1	$\mathcal{N}(0 \text{ pc}, 100 \text{ pc})$	$\ln \delta$	$\mathcal{N}(-0.5, 0.5)$
z_1	$\mathcal{N}(0 \text{ pc}, 50 \text{ pc})$	$\ln \sigma / \text{pc}$	$\mathcal{U}(3.5, 5)$
x_2	$\mathcal{N}(300 \text{ pc}, 100 \text{ pc})$	f	$\mathcal{U}(0.15, 0.4)$
y_2	$\mathcal{N}(1400 \text{ pc}, 100 \text{ pc})$		
z_2	$\mathcal{N}(0 \text{ pc}, 50 \text{ pc})$		

Extended Data Table 2 | Constraints on Radcliffe Wave parameters

Parameter	Median with 95% CI	Parameter	Median with 95% CI
x_0	-910^{+130}_{-150} pc	P	3560^{+500}_{-470} pc
y_0	-860^{+140}_{-130} pc	A	160^{+30}_{-30} pc
z_0	-30^{+80}_{-80} pc	ϕ	$2.89^{+0.51}_{-0.58}$ rad
x_1	-270^{+80}_{-80} pc	γ	$0.50^{+0.27}_{-0.17}$
y_1	-20^{+160}_{-150} pc	δ	$6.68^{+5.48}_{-3.20}$
z_1	-10^{+30}_{-30} pc	σ	62^{+16}_{-13} pc
x_2	290^{+70}_{-70} pc	f	$0.22^{+0.07}_{-0.06}$
y_2	1400^{+170}_{-170} pc		
z_2	30^{+50}_{-50} pc		

CI, credible interval.

Extended Data Table 3 | Physical properties of the Radcliffe Wave

Name	Median with 95% CI
Length	$2.7 \pm 0.2 \text{ kpc}$
Scatter	$60 \pm 15 \text{ pc}$
Amplitude	$160 \pm 30 \text{ pc}$
Mass	$\geq 3 \times 10^6 \text{ M}_{\odot}$

Entanglement of two quantum memories via fibres over dozens of kilometres

<https://doi.org/10.1038/s41586-020-1976-7>

Received: 26 March 2019

Accepted: 12 November 2019

Published online: 12 February 2020

Yong Yu^{1,2,3,6}, Fei Ma^{1,2,3,4,6}, Xi-Yu Luo^{1,2,3}, Bo Jing^{1,2,3}, Peng-Fei Sun^{1,2,3}, Ren-Zhou Fang^{1,2,3}, Chao-Wei Yang^{1,2,3}, Hui Liu^{1,2,3}, Ming-Yang Zheng⁴, Xiu-Ping Xie⁴, Wei-Jun Zhang⁵, Li-Xing You⁵, Zhen Wang⁵, Teng-Yun Chen^{1,2,3}, Qiang Zhang^{1,2,3,4*}, Xiao-Hui Bao^{1,2,3*} & Jian-Wei Pan^{1,2,3*}

A quantum internet that connects remote quantum processors^{1,2} should enable a number of revolutionary applications such as distributed quantum computing. Its realization will rely on entanglement of remote quantum memories over long distances. Despite enormous progress^{3–12}, at present the maximal physical separation achieved between two nodes is 1.3 kilometres¹⁰, and challenges for longer distances remain. Here we demonstrate entanglement of two atomic ensembles in one laboratory via photon transmission through city-scale optical fibres. The atomic ensembles function as quantum memories that store quantum states. We use cavity enhancement to efficiently create atom–photon entanglement^{13–15} and we use quantum frequency conversion¹⁶ to shift the atomic wavelength to telecommunications wavelengths. We realize entanglement over 22 kilometres of field-deployed fibres via two-photon interference^{17,18} and entanglement over 50 kilometres of coiled fibres via single-photon interference¹⁹. Our experiment could be extended to nodes physically separated by similar distances, which would thus form a functional segment of the atomic quantum network, paving the way towards establishing atomic entanglement over many nodes and over much longer distances.

Establishing remote entanglement is a central theme in quantum communication^{1,2,20}. So far, entangled photons have been distributed over long distances both in optical fibres²¹ and in free space with the assistance of satellites²². In spite of this progress, the distribution succeeds only with an extremely low probability owing to severe transmission losses and because photons have to be detected to verify their survival after transmission. Therefore the distribution of entangled photons has not been scalable to longer distances or to multiple nodes^{20,23}. A very promising solution is to prepare separate atom–photon entanglement in two remote nodes and to distribute the photons to an intermediate node for interference^{17,19}. Proper measurement of the photons will project the atoms into a remote entangled state. Although the photons will still undergo transmission losses, the success of remote atomic entanglement will be heralded by the measurement of photons. Therefore, if the atomic states can be stored efficiently for a sufficiently long duration, multiple pairs of heralded atomic entanglement could be further connected efficiently to extend entanglement to longer distances or over multiple quantum nodes through entanglement swapping²³, thus making quantum-internet-based applications feasible^{2,24,25}.

Towards this goal, a great number of experimental investigations have been made with many different matter systems^{23,26–29}, each of which has its own advantages in enabling different capabilities. So far, entanglement of two stationary qubits has been achieved with atomic ensembles^{3,4,6,7}, single atoms⁸, nitrogen vacancy centres^{9,10,12}, quantum

dots¹¹, trapped ions⁵, and so on. Nevertheless, for all systems, the maximum distance between two physically separated nodes remains 1.3 km (ref. ¹⁰). To extend the distance to the city scale, there are three main experimental challenges, which are: to achieve bright (that is, efficient) matter–photon entanglement, to reduce the transmission losses, and to realize stable and high-visibility interference in long fibres. In this Article we combine an atomic-ensemble-based quantum memory with efficient quantum frequency conversion (QFC)¹⁶, and we realize the entanglement of two quantum memories via fibre transmission over dozens of kilometres. We make use of cavity enhancement to create a bright source of atom–photon entanglement. We employ the differential-frequency generation (DFG) process in a periodically poled lithium niobate waveguide (PPLN-WG) chip to shift the single-photon wavelength from the near-infrared part of the spectrum to the telecommunications O band for low-loss transmission in optical fibres. We then make use of a two-photon interference scheme^{17,18} to entangle two atomic ensembles over 22 km of field-deployed fibres. Moreover, we make use of a single-photon interference scheme¹⁹ to entangle two atomic ensembles over 50 km of coiled fibres. Our work can be extended to long-distance separated nodes as a functional segment for atomic quantum networks and quantum repeaters³⁰ and should soon enable repeater-based quantum communications, paving the way towards building large-scale quantum networks over long distances in a scalable way^{1,2}.

¹Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, China. ²Department of Modern Physics, University of Science and Technology of China, Hefei, China. ³CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, China. ⁴Jinan Institute of Quantum Technology, Jinan, China. ⁵State Key Laboratory of Functional Materials for Informatics, Shanghai Institute of Microsystem and Information Technology (SIMIT), Chinese Academy of Sciences, Shanghai, China. ⁶These two authors contributed equally: Yong Yu, Fei Ma. *e-mail: qiangzh@ustc.edu.cn; xhbao@ustc.edu.cn; pan@ustc.edu.cn

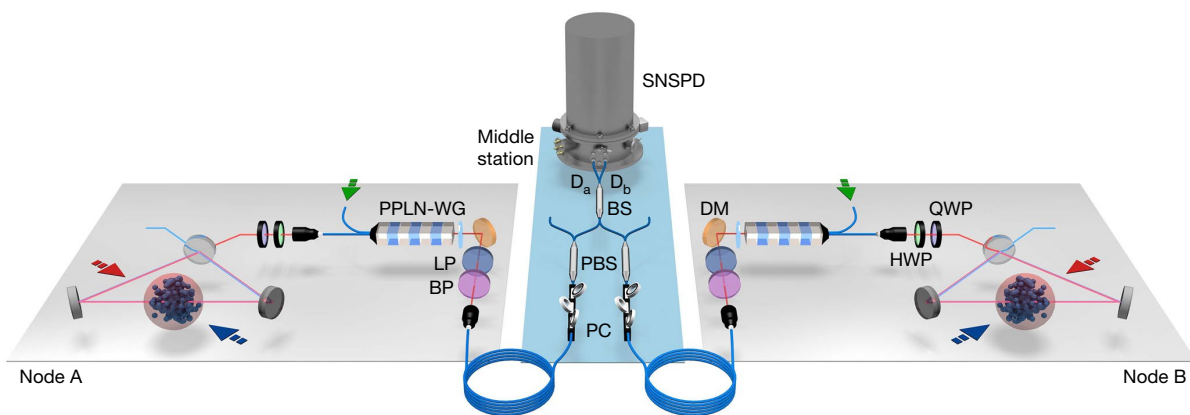


Fig. 1 | Schematic of the remote entanglement generation between atomic ensembles. Two quantum memory nodes (nodes A and B in one laboratory) are linked by fibres to a middle station for photon measurement. In each node, a ^{87}Rb atomic ensemble is placed inside a ring cavity. All atoms are prepared in the ground state at first. We first create a local entanglement between atomic ensemble and a write photon by applying a write pulse (blue arrow). Then the write-out photon is collected along the clockwise (anticlockwise) cavity mode and sent to the QFC module. With the help of a PPLN-WG chip and a 1,950-nm pump laser (green arrow), the 795-nm write-out photon is converted to the telecommunications O band (1,342 nm). The combination of a half-wave-plate (HWP) and a quarter-wave-plate (QWP) improves the coupling with the transverse magnetic polarized mode of the waveguide. After noise filtering,

two write-out photons are transmitted through long fibres, interfered inside a beamsplitter and detected by two superconducting nanowire single-photon detectors (SNSPDs) with efficiencies of about 50% at a dark-count rate of 100 Hz. The effective interference in the middle station heralds two entangled ensembles. Fibre polarization controllers (PCs) and polarization beamsplitters (PBSs) before the interference beamsplitter (BS) are intended to actively compensate polarization drifts in the long fibre. To retrieve the atom state, we apply a read pulse (red arrow) counter-propagating to the write pulse. By phase-matching the spin-wave and cavity enhancement, the atomic state is efficiently retrieved into the anticlockwise (clockwise) mode of the ring cavity. DM refers to dichroic mirror, LP refers to long-pass filter and BP refers to band-pass filter.

Quantum memory with telecommunications interface

Our experiment consists of two similar nodes linked via long-distance fibres, as shown in Fig. 1. In each node, an ensemble of about 10^8 atoms trapped and cooled by laser beams serves as the quantum memory²³. All atoms are initially prepared in the ground state $|g\rangle$. Following the Duan–Lukin–Cirac–Zoller protocol¹⁹, in each trial, a weak write pulse coupling ground-state atoms to the excited state $|e\rangle$ induces a spontaneous Raman-scattered write-out photon together with a collective excitation of the atomic ensemble in a stable state $|s\rangle$ with a small probability χ . The collective excitation can be stored for a long time and later be retrieved on demand as a read-out photon in a phase-matching mode by applying the read pulse, which couples to the transition of $|g\rangle \leftrightarrow |e\rangle$. The write-out and read-out photons are nonclassically correlated. By employing a second Raman-scattering channel $|g\rangle \rightarrow |e\rangle \rightarrow |s'\rangle$, we can create the entanglement between the polarization of the write-out photon and internal state ($|s\rangle$ or $|s'\rangle$) of the atomic ensemble^{15,31}. To further enhance the readout efficiency¹³ and suppress noise from control beams, we build a ring cavity with a finesse (a figure of merit that quantifies the quality of the cavity) of 23.5 around the atomic ensemble. The ring cavity not only enhances the retrieval but also serves as a filter to eliminate the necessity of using external frequency filters to suppress noise³¹.

To create remote atomic entanglement over a long distance, it is crucial that the photons are suitable for low-loss transmission in optical fibres. Therefore we shift the wavelength of the write-out photon from the near-infrared (3.5 dB km^{-1} at 795 nm) to the telecommunications O band (0.3 dB km^{-1} at 1,342 nm) via the DFG process. We make use of reverse-proton-exchange PPLN-WG chips. Optimal coupling efficiency and transmission for the 795-nm signal and the 1,950 nm pump are simultaneously achieved in one chip by an integrated structure consisting of two waveguides (see Supplementary Information). Figure 2a shows that its conversion efficiency is up to $\eta_{\text{conv}} \approx 70\%$ using a 270-mW pump laser. During the conversion, there are three main spectral components of noise: at 1,950 nm, at 975 nm and at 650 nm, which come from the pump laser and its second and third harmonic generation. They are all spectrally far enough away from 1,342 nm that

we can cut them off via the combination of two dichroic mirrors and a long-pass filter with a transition wavelength of 1,150 nm. The pump laser also induces broadband Raman noise, the spectral brightness of which around 1,342 nm is measured to be about 500 Hz nm^{-1} . Thus, we use a bandpass filter (centred at 1,342 nm, with linewidth 5 nm) to confine this noise to approximately 2.5 kHz, which corresponds to a signal-to-noise ratio of $>20:1$, as depicted in Fig. 2a. The filtering process induces only 20% loss, and fibre coupling causes an extra 40% loss. The end-to-end efficiency of our QFC module is $\eta_{\text{QFC}} = 33\%$, which is the highest value for all memory telecommunications quantum interfaces^{32–38} reported so far, to the best of our knowledge. In addition, we perform a Hanbury–Brown–Twiss experiment for the write-out photons with and without QFC, with the results shown in Fig. 2b, which verify that the single-photon quality is well preserved during QFC.

Entanglement over 22 km of field fibres

We first perform a two-node experiment via two-photon interference (TPI)¹⁸. In each node, we create entanglement between polarization of the write-out photon and the internal state of the collective excitation via a ‘double- Λ ’-shaped level scheme (see Supplementary Information for level details). The entangled state can be expressed as $(|\uparrow\odot\rangle + |\downarrow\odot\rangle)/\sqrt{2}$, where $|\uparrow\rangle$ or $|\downarrow\rangle$ denotes an atomic excitation in $|s\rangle$ or $|s'\rangle$ respectively, and $|\odot\rangle$ and $|\ominus\rangle$ denote polarization of the write-out photon. To characterize the atom–photon entanglement, we perform quantum state tomography, with the result shown in Fig. 3. We obtain a fidelity of 0.930(6) for node A and of 0.933(6) for node B when $\chi = 0.019$. The two nodes are located in one laboratory on the east campus of the University of Science and Technology of China ($31^\circ 50' 6.96'' \text{N}$, $117^\circ 15' 52.07'' \text{E}$), as shown in Fig. 4a. Once the polarization entanglement is ready, the write-out photon is converted locally by QFC into the telecommunications band. Two photons from different nodes are transmitted along two parallel field-deployed commercial fibre channels (11 km per channel) from the University of Science and Technology of China to the Hefei Software Park ($31^\circ 51' 6.01'' \text{N}$, $117^\circ 11' 54.72'' \text{E}$), as shown in Fig. 4a. Over there, we perform a Bell-state measurement by detecting two photons simultaneously with superconducting

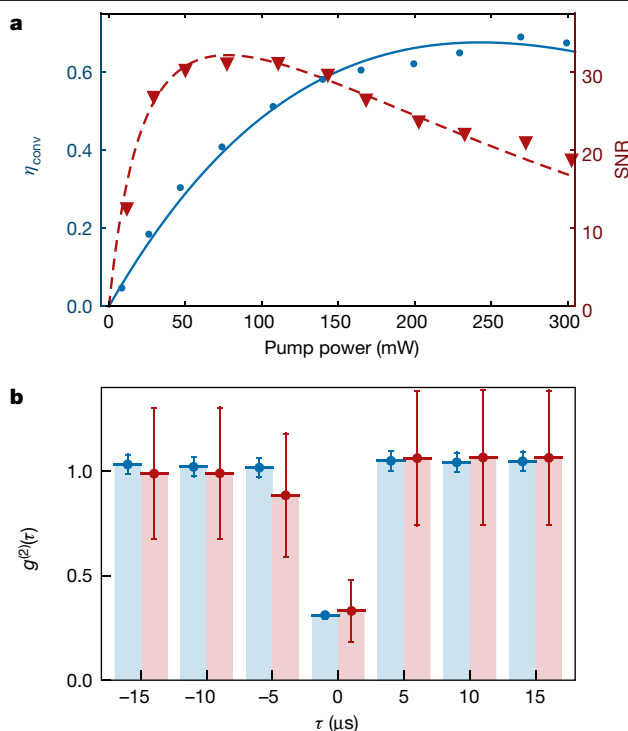


Fig. 2 | Performance of the telecommunications interface. a, The conversion efficiency η_{conv} and the signal-to-noise ratio (SNR) vary as a function of pump laser power. Blue dots refer to the overall conversion efficiency of the PPLN-WG chip, and red triangles refer to the signal-to-noise ratio at probability $\chi = 0.015$. **b**, Measurement of the second-order correlation function $g^{(2)}$ (results of the Hanbury–Brown–Twiss experiment) with (red) and without (blue) QFC at $\chi = 0.057$. The write-out photons are only measured if the corresponding read-out photon is detected. The error bars represent one standard deviation.

nanowire single photon detectors. A successful Bell-state measurement result projects, in a heralded way, the two atomic ensembles into a maximally entangled state:

$$|\Psi^{\pm}\rangle_{\text{TPI}} = \frac{1}{\sqrt{2}}(|\uparrow\rangle_A |\downarrow\rangle_B \pm |\downarrow\rangle_A |\uparrow\rangle_B) \quad (1)$$

with a internal sign determined by the measurement outcome of the Bell-state measurement.

The strong polarization dependence of DFG in the PPLN-WG makes it difficult to perform QFC directly for a polarization encoded photon. In this experiment, we transform the polarization encoding into time-bin encoding³⁹ and let the two photonic modes pass through the QFC

module in sequence with the same polarization. As shown in Fig. 4b, the transformation is realized through an AMZI and a fast Pockels cell which erases the polarization distinguishability. For the time-bin encoding, it is crucial that the two modes have a stable relative phase shift, which is realized via active stabilization of the two AMZIs. Moreover, the transformation into time-bin encoding offers an additional advantage of robustness in long-distance transmission in fibres.

Before long-fibre experiments, we characterize the atom–atom entanglement locally without QFC. For the measurement of the atomic qubits, we first apply Raman rotations⁴⁰, then we retrieve the excitations into read-out photons and measure the polarization³¹. Measurement in an arbitrary basis is realized by configuring the Raman pulses. Figure 5a shows the measured fidelity averaged for $|\Psi^{\pm}\rangle$ as a function of χ . At $\chi \approx 2\%$, we get $\mathcal{F} = 0.798 \pm 0.063$ for $|\Psi^{\pm}\rangle$ and 0.829 ± 0.036 for $|\Psi^{\pm}\rangle$, respectively, which are in good agreement with the theoretical estimation. Furthermore, the fidelity is almost independent of χ , after subtracting the accidental coincidences that are mainly due to high-order excitations in the Raman scattering process.

The field-deployed long fibre ($L = 22$ km) induces 8 dB of attenuation. Besides, the long fibre leads to random rotations of polarization. To optimize the indistinguishability, we apply polarization filtering for the photons after long-fibre transmission before the Bell-state measurement. In addition, to get a high filtering efficiency, we perform active polarization compensation by replacing the manual polarization controllers in Fig. 1 with electric polarization controllers and minimizing the reflections of the filtering polarization beamsplitters. We get an average efficiency of 98%, as shown in Fig. 4c. To reduce the background noise in the fibre channels, we carefully cover all the fusion points and get an average background noise of about 280 Hz (including dark counts of the detector). In the long-fibre case, to increase the count rate, we set the excitation probability to $\chi = 0.038$ and perform entanglement verification in a delayed-choice fashion⁴¹. The measured visibility in the $|\uparrow\rangle/|\downarrow\rangle$ basis is $V_1 = 0.684 \pm 0.075$ for $|\Psi^{\pm}\rangle$ and $V_1 = 0.635 \pm 0.075$ for $|\Psi^{\pm}\rangle$. Adjusting the Raman pulse delay δt , we could observe a sinusoidal oscillation in the $|\uparrow\rangle \pm |\downarrow\rangle$ basis as shown in Fig. 5b with a visibility of $V_2 = 0.574 \pm 0.064$ for $|\Psi^{\pm}\rangle$ and $V_2 = 0.647 \pm 0.066$ for $|\Psi^{\pm}\rangle$. By assuming a similar visibility in the $|\uparrow\rangle \pm |\downarrow\rangle$ basis, the entanglement fidelity can be estimated as⁴² $\mathcal{F} \approx \frac{1}{4}(1 + V_1 + 2V_2) = 0.708 \pm 0.037$ for $|\Psi^{\pm}\rangle$, and 0.732 ± 0.038 for $|\Psi^{\pm}\rangle$, which greatly exceed the bound of $\mathcal{F} > 0.5$ required to witness entanglement for a Bell state. The measured heralding rate is $P_{\text{her}} = 1.46 \times 10^{-6}$, half of which is due to double-excitation events from a single node. Thus the entangling probability is estimated to be $P_{\text{ent}} \approx P_{\text{her}}/2 = 0.73 \times 10^{-6}$.

Entanglement over 50 km of coiled fibres

The entangling probability in the TPI experiment¹⁸ is low since it scales as χ^2 and $\eta_{L/2}^2$, where $\eta_{L/2}$ is the overall optical efficiency from one node

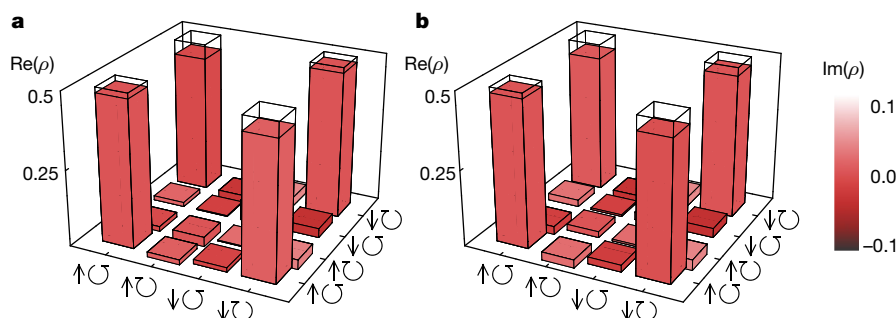


Fig. 3 | Tomography of the atom–photon entanglement. a, b, The reconstructed density matrix between the write-out photon and the atomic spin-wave in node A (a) and in node B (b). In each element of the matrix, the

height of the bar represents its real part, $\text{Re}(\rho)$, and the colour represents its imaginary part, $\text{Im}(\rho)$. The transparent bars indicate the ideal density matrix of the maximally entangled state.

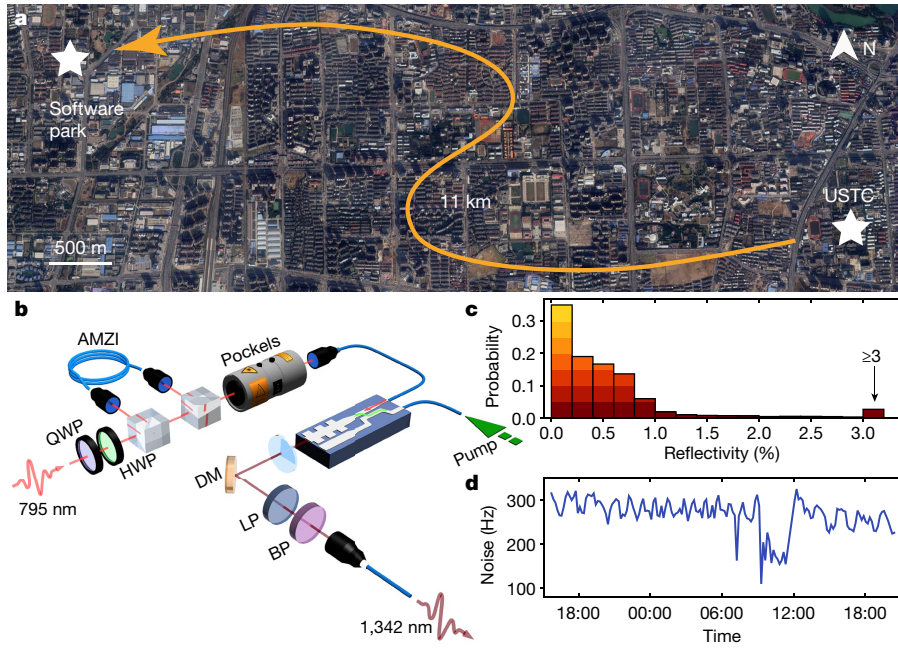


Fig. 4 | Entanglement over field fibres. **a**, Bird's-eye view of the remote entanglement experiment over the field fibre. Two quantum nodes are located at the University of Science and Technology of China (USTC). Telecommunications photons from two nodes are transmitted through two parallel field-deployed fibres to the middle station located at the Hefei Software Park. Each fibre is 11 km long and has an 4-dB attenuation for the 1,342-nm photon. (Map data from Google, Maxar Technologies.) **b**, Setup for polarization photon QFC. Two polarization beamsplitters and a coiled polarization-maintaining delay fibre constitute an asymmetric Mach-Zender

interferometer (AMZI). Two orthogonal polarization components ($|0\rangle$ and $|1\rangle$) of the 795-nm photon are separated in the time domain after the AMZI, and the polarization information is actively erased by a Pockels cell. Then the time-bin encoded photon is sent to the QFC module. **c**, Probability distribution of the reflectivity for the polarization-filtering polarization beamsplitters (shown in Fig. 1 after long fibres), with active compensation. The data shown was recorded once per second and accumulated over 24 hours. **d**, Background noise in the superconducting nanowire single-photon detector over 24 hours.

to the Bell-state measurement. In contrast, a single-photon interference (SPI) scheme⁴⁹ gives an entangling probability that scales linearly as a function of χ and $\eta_{L/2}$. Thus, targeting a much higher entangling probability, we perform another two-node experiment via SPI. As shown in Fig. 1, two pairs of Fock-state entanglement are created at nodes A and B, respectively, in the form of $|0\rangle_p|0\rangle_a + \sqrt{\chi}|1\rangle_p|1\rangle_a$, where 0 and 1 represent the number of photons or atomic excitations. The frequency-converted photons from both nodes are then transmitted along a long fibre, later combined through a fibre beamsplitter to perform SPI and eliminate its 'which way' (that is, which fibre path the photon travels through) information, finally detected with superconducting nanowire single photon detectors. A click from detectors D_a or D_b (shown in Fig. 1) heralds that two ensembles are mapped into a maximally entangled state:

$$|\Psi^\pm\rangle_{\text{SPI}} = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B \pm e^{i\Delta\phi}|1\rangle_A|0\rangle_B) \quad (2)$$

where $\Delta\phi$ is the accumulated phase difference between two fibre channels. To keep $\Delta\phi$ in equation (2) constant, we harness an intermittent phase-locking loop in situ during every experimental interval to eliminate phase drift (see Supplementary Information).

To verify the Fock-state atomic entanglement, we follow a protocol introduced in ref. ⁴. The degree of entanglement is quantified in terms of concurrence C , which is a monotone function of entanglement and goes from 0 for a separable state to 1 for a maximally entangled state. Its definition is $C = \max(0, 2|d| - 2\sqrt{p_{00}p_{11}})/P$, where p_{ij} is the probability of having i excitations in ensemble A and having j excitations in ensemble B, $P = p_{00} + p_{01} + p_{10} + p_{11}$, $d \approx V(p_{01} + p_{10})/2$, and V is the interference visibility of the single-excitation states. The excitation statistics of p_{ij} can be measured directly via photon counting of the two read-out modes and applying loss calibration. To measure the interference

visibility V , we add a relative phase θ between two read-out modes and mix them via a beamsplitter. Along with the scan of θ , counts in two output modes vary as a sinusoidal function of θ as shown in Fig. 6, and thus we could deduce $V = V_\theta$. For the short-fibre case ($L = 10$ m) without QFC, at $\chi = 0.015$, we get a concurrence of $C = 0.677 \pm 0.012$ for $|\Psi^+\rangle$ and $C = 0.711 \pm 0.012$ for $|\Psi^-\rangle$. In this condition, the entangling probability in one trial is $P_{\text{ent}} = 0.014$, which is the highest probability of heralded remote entanglement creation to the best of our knowledge.

In the long-fibre ($L = 10$ km and 50 km) cases, we add the QFC module. The phase noise during long-fibre transmission fluctuates faster⁴³ than the capable band of intermittent phase-locking. Hence we additionally insert an auxiliary continuous 1,550-nm laser beam to uninterruptedly monitor phase fluctuation and actively stabilize it (see Supplementary Information). Measured results for the read-out photon interference at different fibre lengths are shown in Fig. 6. By fitting the sinusoidal oscillations and measuring the excitation statistics, we get a concurrence result of $C = 0.428 \pm 0.013$ at $L = 10$ km and $C = 0.407 \pm 0.008$ at $L = 50$ km for $|\Psi^+\rangle$. For $|\Psi^-\rangle$, the results are $C = 0.416 \pm 0.008$ at $L = 10$ km and $C = 0.348 \pm 0.011$ at $L = 50$ km. Degradation of concurrence in comparison with the case of short fibre without QFC is mainly due to the remaining noise after phase stabilization (see Supplementary Information), which can be greatly improved by optimizing the feedback loop. The measured heralded entangling probability is $P_{\text{ent}} = 1.57 \times 10^{-3}$ for the 10-km fibre and $P_{\text{ent}} = 3.85 \times 10^{-4}$ for the 50-km fibre, which correspond to an entanglement creation time of $T_{\text{ent}} = 32$ ms and $T_{\text{ent}} = 0.65$ s, respectively.

Discussion and outlook

We have experimentally demonstrated two feasible ways to entangle two quantum memories via long-distance photon transmission in optical fibres. We summarize key parameters and results in Table 1.

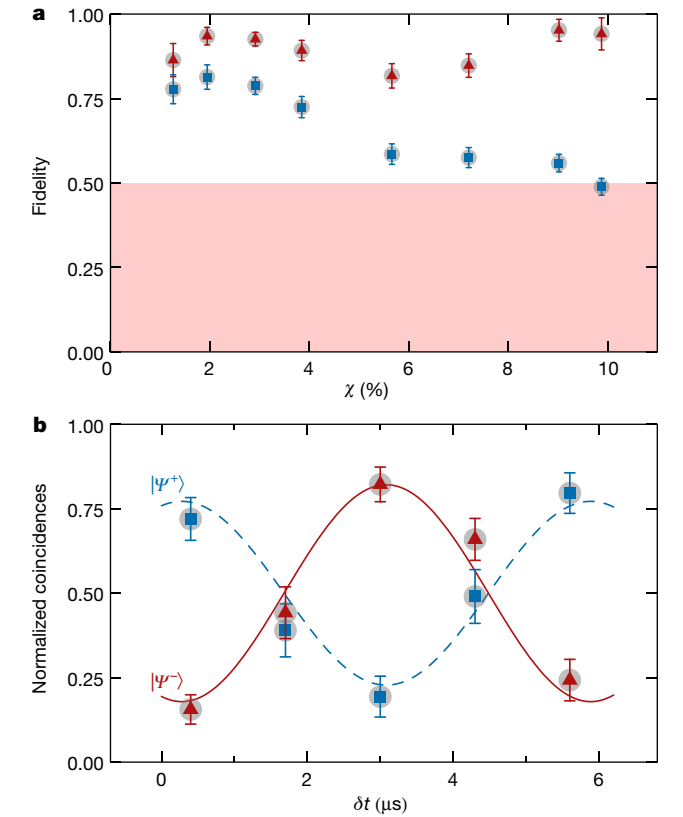


Fig. 5 | Characterization of the remote entanglement via TPI. **a**, Average fidelity of the remote entanglement $|\Psi^+\rangle$ generated locally as a function of χ . Blue squares refer to the measurement result. Red triangles show the corrected results through deduction of accidental coincidences (see Supplementary Information). The error bars represent one standard deviation. Pink shading indicates where fidelity is not sufficient to claim entanglement. **b**, Coincidences measured in the $|\pm\rangle = |\uparrow\rangle \pm |\downarrow\rangle$ basis for the two atomic qubits, normalized by the total coincidence of all combinations. The Raman pulse in node A is applied slightly later than in node B with an offset of δt , which induces a linearly changing phase in Ψ^+ and results in the observed oscillations. Parallel correlations ($|\uparrow\rangle\langle\uparrow|$ or $|\downarrow\rangle\langle\downarrow|$) of $|\Psi^+\rangle$ (blue squares) and $|\Psi^-\rangle$ (red triangles) are shown. Solid red and dashed blue lines correspond to the fitting results. The 5.4-s oscillation period agrees with Zeeman splitting between $|\uparrow\rangle$ and $|\downarrow\rangle$. This plot is based on 2.9×10^4 heralding events during a total measurement time of 487 hours over a period of 30 days. The error bars represent one standard deviation.

Even though the fibre distance of the SPI experiment is much longer than in the TPI experiment, the SPI scheme offers a much higher probability of entanglement creation, because only a single photon passing through half of the whole link is detected. In contrast, the TPI scheme requires the detection of two photons passing through the whole link. For the extension to physically separated nodes over long distances, the TPI scheme is straightforward, merely requiring that photons be indistinguishable. However, the extension of the SPI experiment requires more effort because the scheme is phase-sensitive. According to our analysis in the Supplementary Information, the main difficulty is to achieve phase correlation of remote independent control lasers. We have performed a preliminary test with two lasers locked independently to two ultrastable cavities, which shows that phase correlation can be built and stable for a duration that is long enough to generate remote entanglement (see Supplementary Information for details). Thus it is also feasible to extend our SPI experiment to long-distance separated nodes.

The quantum link efficiency¹² η_{link} , defined as the ratio of memory lifetime over entanglement generation time, is also an important figure of merit for two-node experiments. In our current work, decoherence

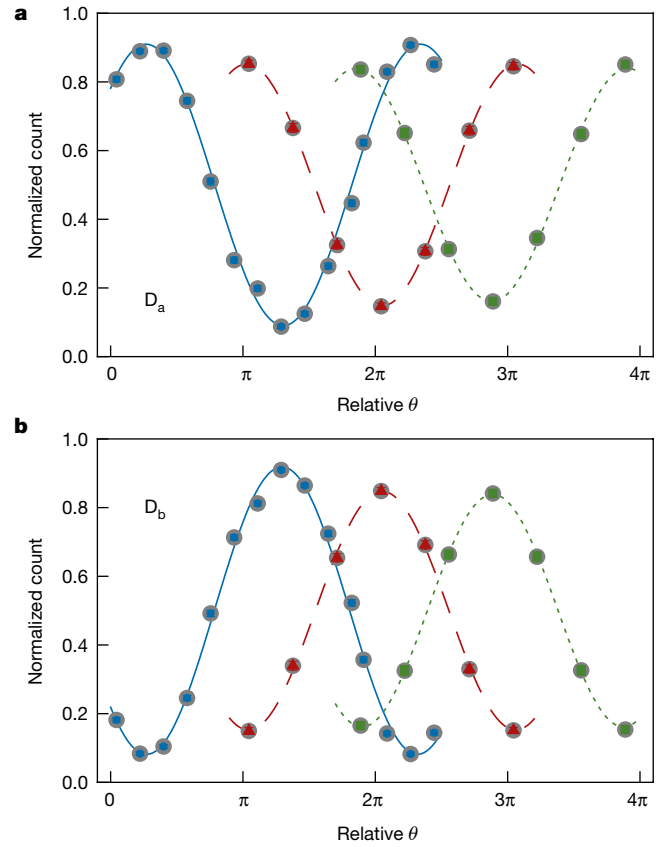


Fig. 6 | Characterization of the remote entanglement via SPI. When the atomic modes are retrieved as optical modes for interference, the photon count in one output mode of the fibre beamsplitter oscillates as a function of the relative phase θ between the two optical modes, normalized by the total count of two output modes. Detector D_a heralded events are shown in **a**; detector D_b heralded events are shown in **b**. Blue squares, red triangles and green dots refer to $L = 10$ m, 10 km and 50 km separately. Sinusoids with corresponding colour (solid, dashed and dotted) show the fitting results. The 50-km result is based on 1.7×10^5 heralding events during a total measurement time of 6 hours over a period of 2 days. The error bars represent one standard deviation.

due to atom motions results in a memory lifetime of about 70 μ s, which is much smaller than the entanglement generation time. According to our previous work on a very similar setup⁴⁴, applying a three-dimensional optical lattice can improve the lifetime to the sub-second

Table 1 | Comparison of two-node experiments

Experiment	TPI (this work)	SPI (this work)	NV (2015; ref. ¹⁰)
Physical separation	0.6 m	0.6 m	1.3 km
Overall fibre length, L	22 km	50 km	1.7 km
Entanglement probability, P_{ent}	0.73×10^{-6}	3.85×10^{-4}	6.4×10^{-9}
Entanglement quality	$\mathcal{F} = 0.720 \pm 0.027$	$\mathcal{C} = 0.378 \pm 0.007$	$\mathcal{F} = 0.92 \pm 0.03$
Entanglement creation time, T_{ent}	150 s	0.65 s	1.3×10^3 s
Quantum link efficiency, η_{link} (ref. ¹²)	1.45×10^{-3}	0.34	4.6×10^{-4}
Assumed memory lifetime, τ_m	0.22 s (ref. ⁴⁴)	0.22 s (ref. ⁴⁴)	0.6 s (refs. ^{12,51})

In the long-fibre case, the propagation delay results in a maximal repetition rate of $R_{\text{rep}} = C/L$, where $C = 2 \times 10^8$ m s⁻¹ is the speed of light in the fibre. Thus the heralded entanglement creation time is estimated as $T_{\text{ent}} = (R_{\text{rep}} P_{\text{ent}})^{-1}$. For the estimation of $\eta_{\text{link}} = \tau_m / T_{\text{ent}}$, we make use of state-of-the-art lifetime results, as listed in the last row. We chose the 1.3-km nitrogen vacancy (NV) experiment¹⁰ for comparison, because it is the only previous two-node experiment that has a fibre length in the kilometre regime.

regime ($\tau_m = 0.22$ s), on the basis of which the quantum link efficiency is estimated to be $\eta_{\text{link}} = 0.34$ for SPI and $\eta_{\text{link}} = 2.9 \times 10^{-3}$ for TPI. For further improvement of η_{link} , it is crucial to increase the entanglement generation rate. For example, one may use Rydberg blockade to inhibit the high-order excitations during atom–photon entanglement preparation, and make the preparation process deterministic^{45,46}. One can also make use of the multiplexing technique^{47–50} to prepare multiplexed atom–photon entanglement. Shifting the wavelength to the telecommunications C band, optimizing the coupling efficiencies and using better detectors will also greatly increase the remote entanglement rate.

Extending these experiments to nodes separated by much longer distances will enable us to perform advanced quantum information tasks, such as efficient quantum teleportation over long distances. By incorporating more quantum memories, our experiment may be extended to entangle multiple quantum memories over long distances via multi-photon interference³¹. One may also create two pairs of remote atomic entanglement over two sub-links and extend the distance of atomic entanglement via entanglement swapping, following the quantum repeater scheme³⁰. Concatenating this process could extend the distance sufficiently to beat the limit of direct transmission²³.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1976-7>.

- Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
- Wehner, S., Elkouss, D. & Hanson, R. Quantum internet: a vision for the road ahead. *Science* **362**, eaam9288 (2018).
- Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
- Chou, C. W. et al. Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438**, 828–832 (2005).
- Moehring, D. L. et al. Entanglement of single-atom quantum bits at a distance. *Nature* **449**, 68–71 (2007).
- Chou, C.-W. et al. Functional quantum nodes for entanglement distribution over scalable quantum networks. *Science* **316**, 1316–1320 (2007).
- Yuan, Z.-S. et al. Experimental demonstration of a BDCZ quantum repeater node. *Nature* **454**, 1098–1101 (2008).
- Hofmann, J. et al. Heralded entanglement between widely separated atoms. *Science* **337**, 72–75 (2012).
- Bernien, H. et al. Heralded entanglement between solid-state qubits separated by three metres. *Nature* **497**, 86–90 (2013).
- Hensen, B. et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
- Delteil, A. et al. Generation of heralded entanglement between distant hole spins. *Nat. Phys.* **12**, 218–223 (2016).
- Humphreys, P. C. et al. Deterministic delivery of remote entanglement on a quantum network. *Nature* **558**, 268–273 (2018).
- Simon, J., Tanji, H., Thompson, J. K. & Vuletić, V. Interfacing collective atomic excitations and single photons. *Phys. Rev. Lett.* **98**, 183601 (2007).
- Bao, X.-H. et al. Efficient and long-lived quantum memory with cold atoms inside a ring cavity. *Nat. Phys.* **8**, 517–521 (2012).
- Yang, S.-J. et al. Highly retrievable spin-wave–photon entanglement source. *Phys. Rev. Lett.* **114**, 210501 (2015).
- Kumar, P. Quantum frequency conversion. *Opt. Lett.* **15**, 1476–1478 (1990).
- Simon, C. & Irvine, W. T. M. Robust long-distance entanglement and a loophole-free Bell test with ions and photons. *Phys. Rev. Lett.* **91**, 110405 (2003).
- Zhao, B., Chen, Z.-B., Chen, Y.-A., Schmiedmayer, J. & Pan, J.-W. Robust creation of entanglement between remote memory qubits. *Phys. Rev. Lett.* **98**, 240502 (2007).
- Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Yuan, Z.-S. et al. Entangled photons and quantum communication. *Phys. Rep.* **497**, 1–40 (2010).
- Inagaki, T., Matsuda, N., Tadanaga, O., Asobe, M. & Takesue, H. Entanglement distribution over 300 km of fiber. *Opt. Express* **21**, 23241–23249 (2013).
- Yin, J. et al. Satellite-based entanglement distribution over 1200 kilometers. *Science* **356**, 1140–1144 (2017).
- Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. *Rev. Mod. Phys.* **83**, 33–80 (2011).
- Gottesman, D., Jennewein, T. & Croke, S. Longer-baseline telescopes using quantum repeaters. *Phys. Rev. Lett.* **109**, 070503 (2012).
- Kómár, P. et al. A quantum network of clocks. *Nat. Phys.* **10**, 582–587 (2014).
- Tittel, W. et al. Photon-echo quantum memory in solid state systems. *Laser Photon. Rev.* **4**, 244–267 (2010).
- Duan, L.-M. & Monroe, C. Quantum networks with trapped ions. *Rev. Mod. Phys.* **82**, 1209–1224 (2010).
- Reiserer, A. & Rempe, G. Cavity-based quantum networks with single atoms and optical photons. *Rev. Mod. Phys.* **87**, 1379–1418 (2015).
- Aharonovich, I., Englund, D. & Toth, M. Solid-state single-photon emitters. *Nat. Photon.* **10**, 631–641 (2016).
- Briegleb, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Jing, B. et al. Entanglement of three quantum memories via interference of three single photons. *Nat. Photon.* **13**, 210–213 (2019).
- Radnaev, A. G. et al. A quantum memory with telecom-wavelength conversion. *Nat. Phys.* **6**, 894–899 (2010).
- De Greve, K. et al. Quantum-dot spin–photon entanglement via frequency downconversion to telecom wavelength. *Nature* **491**, 421–425 (2012).
- Maring, N. et al. Photonic quantum state transfer between a cold atomic gas and a crystal. *Nature* **551**, 485–488 (2017).
- Bock, M. et al. High-fidelity entanglement between a trapped ion and a telecom photon via quantum frequency conversion. *Nat. Commun.* **9**, 1998 (2018).
- Ikuta, R. et al. Polarization insensitive frequency conversion for an atom–photon entanglement distribution via a telecom network. *Nat. Commun.* **9**, 1997 (2018).
- Walker, T. et al. Long-distance single photon transmission from a trapped ion via quantum frequency conversion. *Phys. Rev. Lett.* **120**, 203601 (2018).
- Dréau, A., Tchegoratcheva, A., Mahdaoui, A. E., Bonato, C. & Hanson, R. Quantum frequency conversion of single photons from a nitrogen-vacancy center in diamond to telecommunication wavelengths. *Phys. Rev. Appl.* **9**, 064031 (2018).
- Farrera, P., Heinze, G. & de Riedmatten, H. Entanglement between a photonic time-bin qubit and a collective atomic spin excitation. *Phys. Rev. Lett.* **120**, 100501 (2018).
- Jiang, Y., Rui, J., Bao, X.-H. & Pan, J.-W. Dynamical zeroing of spin-wave momentum to suppress motional dephasing in an atomic-ensemble quantum memory. *Phys. Rev. A* **93**, 063819 (2016).
- Ma, X.-s. et al. Experimental delayed-choice entanglement swapping. *Nat. Phys.* **8**, 479–484 (2012).
- Gühne, O. & Tóth, G. Entanglement detection. *Phys. Rep.* **474**, 1–75 (2009).
- Minář, J., De Riedmatten, H., Simon, C., Zbinden, H. & Gisin, N. Phase-noise measurements in long-fiber interferometers for quantum-repeater applications. *Phys. Rev. A* **77**, 052325 (2008).
- Yang, S.-J., Wang, X.-J., Bao, X.-H. & Pan, J.-W. An efficient quantum light–matter interface with sub-second lifetime. *Nat. Photon.* **10**, 381–384 (2016).
- Li, L., Dudin, Y. O. & Kuzmich, A. Entanglement between light and an optical atomic excitation. *Nature* **498**, 466–469 (2013).
- Li, J. et al. Hong-Ou-Mandel interference between two deterministic collective excitations in an atomic ensemble. *Phys. Rev. Lett.* **117**, 180501 (2016).
- Collins, O. A., Jenkins, S. D., Kuzmich, A. & Kennedy, T. A. B. Multiplexed memory-insensitive quantum repeaters. *Phys. Rev. Lett.* **98**, 060502 (2007).
- Pu, Y.-F. et al. Experimental realization of a multiplexed quantum memory with 225 individually accessible memory cells. *Nat. Commun.* **8**, 15359 (2017).
- Tian, L. et al. Spatial multiplexing of atom-photon entanglement sources using feedforward control and switching networks. *Phys. Rev. Lett.* **119**, 130505 (2017).
- Parniak, M. et al. Wavevector multiplexed atomic quantum memory via spatially-resolved single-photon detection. *Nat. Commun.* **8**, 2140 (2017).
- Bar-Gill, N., Pham, L., Jarmola, A., Budker, D. & Walsworth, R. Solid-state electronic spin coherence time approaching one second. *Nat. Commun.* **4**, 1743 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

Time sequences

Our experiment runs periodically, with each period being composed of an atomic loading phase and an entangling phase. Each loading phase takes 18 ms, during which we reload and cool the atoms and perform active phase-locking. In the entangling phase lasting 2 ms, we repeat entangling trials. For the SPI experiment, each trial lasts 5 μ s including 3 μ s for optical pumping and 2 μ s for the write and read process. For the TPI experiment, each trial lasts 11 μ s including 3 μ s for optical pumping and 8 μ s for the write and read process. The storage duration (relative delay of the read pulse in comparison with the write pulse) is 7 μ s for the TPI experiment and 100 ns for the SPI experiment.

Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements This work was supported by the National Key R&D Program of China (2017YFA0303902, 2018YFB0504300, 2017YFA0304000), the Anhui Initiative in Quantum Information Technologies, the National Natural Science Foundation of China and the Chinese Academy of Sciences. We thank QuantumCTek for providing the field-deployed fibres.

Author contributions X.-H.B. and J.-W.P. conceived the research. Q.Z., X.-H.B. and J.-W.P. designed the experiment. YY., X.-Y.L., B.J., P.-F.S., R.-Z.F., C.-W.Y. and X.-H.B. carried out the experiment with assistance from all other authors. F.M., M.-Y.Z., X.-P.X. and Q.Z. built the QFC module. W.-J.Z., L.-X.Y. and Z.W. fabricated the superconducting nanowire single-photon detectors. YY., Q.Z., X.-H.B. and J.-W.P. analysed the data and wrote the paper with input from all other authors.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1976-7>.

Correspondence and requests for materials should be addressed to Q.Z., X.-H.B. and J.-W.P.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Electrically pumped topological laser with valley edge modes

<https://doi.org/10.1038/s41586-020-1981-x>

Received: 24 April 2019

Accepted: 9 December 2019

Published online: 12 February 2020

Yongquan Zeng¹, Udvash Chattopadhyay², Bofeng Zhu², Bo Qiang^{1,2}, Jinghao Li¹, Yuhao Jin¹, Lianhe Li³, Alexander Giles Davies³, Edmund Harold Linfield³, Baile Zhang^{2*}, Yidong Chong^{2*} & Qi Jie Wang^{1,2*}

Quantum cascade lasers are compact, electrically pumped light sources in the technologically important mid-infrared and terahertz region of the electromagnetic spectrum^{1,2}. Recently, the concept of topology³ has been expanded from condensed matter physics into photonics⁴, giving rise to a new type of lasing^{5–8} using topologically protected photonic modes that can efficiently bypass corners and defects⁴. Previous demonstrations of topological lasers have required an external laser source for optical pumping and have operated in the conventional optical frequency regime^{5–8}. Here we demonstrate an electrically pumped terahertz quantum cascade laser based on topologically protected valley edge states^{9–11}. Unlike topological lasers that rely on large-scale features to impart topological protection, our compact design makes use of the valley degree of freedom in photonic crystals^{10,11}, analogous to two-dimensional gapped valleytronic materials¹². Lasing with regularly spaced emission peaks occurs in a sharp-cornered triangular cavity, even if perturbations are introduced into the underlying structure, owing to the existence of topologically protected valley edge states that circulate around the cavity without experiencing localization. We probe the properties of the topological lasing modes by adding different outcouplers to the topological cavity. The laser based on valley edge states may open routes to the practical use of topological protection in electrically driven laser sources.

Quantum cascade lasers (QCLs) are electrically pumped semiconductor lasers based on intersubband electron transitions within multiple quantum wells in semiconductors^{1,2}. They are among the most important sources of mid-infrared and terahertz (THz) radiation owing to their compactness, electrical pumping performance and high efficiency¹³. Their practical applications include telecommunication¹⁴, THz signal processing¹⁵, imaging¹⁶, sensing and spectroscopy. As with any laser, the emission characteristics of a THz QCL depend on the design of the photonic cavity and are generally strongly affected by the cavity shape^{17,18}. One promising design is the use of topological edge states, which form running-wave modes that are robust against perturbations to the underlying structure^{5–8} and can efficiently bypass defects (which may arise during fabrication and packaging) and sharp corners. Unlike conventional waves, topological edge states resist the formation of localized standing-wave modes, which is helpful for suppressing the spatial hole-burning effect^{19,20}. This is a particularly important consideration for QCLs because their gain recovery processes are faster than the carrier diffusion, unlike in traditional semiconductor lasers²¹.

Topological edge states arise at the interface between spatial domains that have topologically distinct bandstructures³. There have been substantial efforts to implement such states in photonics, motivated by potential applications in robust optical delay lines²²,

amplifiers²³ and other devices^{24,25}. Topological lasers have been realized in one-dimensional (1D) Su–Schrieffer–Heeger (SSH)-like systems^{26,27}, whose edge states act as high-*Q* (quality factor) nanocavity modes that lase under suitable gain. However, the edge states of 1D lattices do not support protected transport. For two-dimensional (2D) lattices, realizing photonic topological edge states typically requires some means of effective breaking of time-reversal (*T*) symmetry to avoid the need to use magnetic materials⁴. For example, a recent demonstration of 2D topological lasing^{5–7} used an array of ring resonators in which the clockwise (CW) or counterclockwise (CCW) circulation of light in the resonators acts as a photonic pseudospin; staggered inter-resonator couplings generate an effective magnetic field and hence a *T*-broken band structure with non-trivial topology for each pseudospin²². This design requires large-scale structural features (for example ring resonators) far exceeding the operating wavelength.

Valley photonic crystals (VPCs)^{10,11} are photonic analogues of 2D valleytronic materials¹² that host topological edge states protected by a valley degree of freedom established by the underlying lattice symmetry. They have been demonstrated in several photonic crystal geometries^{28–30}, and similar valley-protected edge states have been realized in sonic crystals³¹. In 2D materials, the valley degree of freedom can function similarly to spin in a spintronic device but does not require strong

¹Centre for OptoElectronics and Biophotonics, School of Electrical and Electronic Engineering & The Photonics Institute, Nanyang Technological University, Singapore, Singapore. ²Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore. ³School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK. *e-mail: blzhang@ntu.edu.sg; yidong@ntu.edu.sg; qjwang@ntu.edu.sg

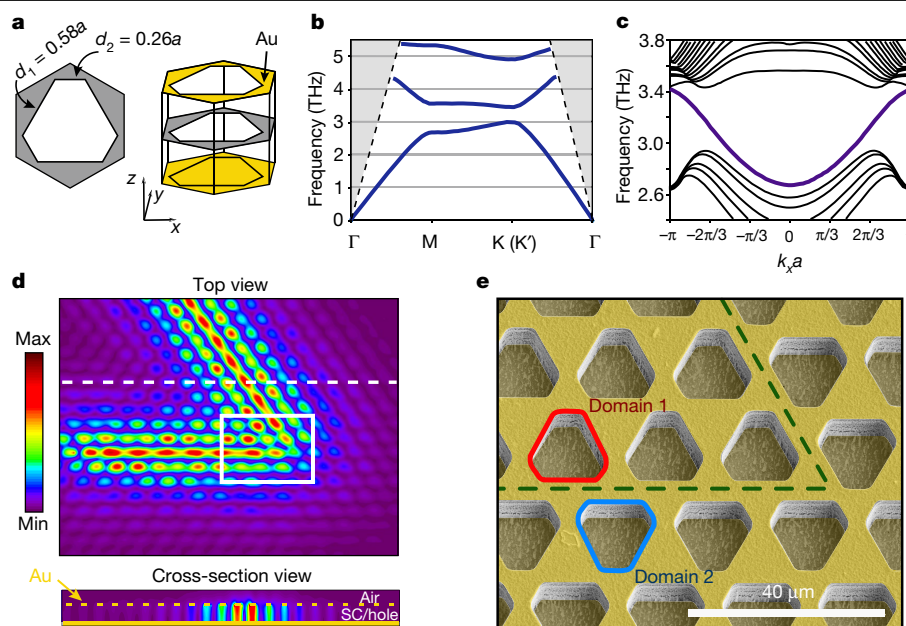


Fig. 1 | Design of a terahertz quantum cascade laser with topologically protected valley edge modes. **a**, Each unit cell of the valley photonic crystal contains a quasi-hexagonal hole perforated through the top metal and the semiconductor layer in a metal–semiconductor–metal structure. The lattice period is $a = 19.5 \mu\text{m}$. **b**, Band structure calculated by 3D finite-element simulation. **c**, Projected band diagram for a supercell representing a straight domain wall separating two domains with opposite hole orientations, with 10

quasi-hexagonal holes on each side. **d**, Simulated electric field distribution ($|E_z|$) (top view and cross-section view) of a transmission mode in a topological waveguide with a 120° corner. The white dashed line indicates the position of the cross-section view. SC, semiconductor. **e**, SEM image of a portion of the fabricated topological waveguide near the corner, corresponding to the area enclosed by a white rectangle in **d**. Domains 1 and 2 have opposite orientations and thus opposite valley Chern numbers.

spin–orbit coupling¹². Likewise, VPCs can provide robust light transport in highly compact structures with periodicity of the order of the wavelength^{10,11}, without the need for magnetic materials or the complex construction of photonic pseudospins. They are therefore promising for the implementation of compact topological photonic crystal lasers.

We have realized electrically pumped THz QCLs using the topological edge states of a VPC. Lasing is achieved using a topological waveguide that forms a triangular loop, very different from conventional smoothly shaped optical cavities. Despite the sharp corners of the cavity, we find that the lasing spectrum exhibits robust regularly spaced emission peaks, a feature that persists under disturbances including a point outcoupling defect along an arm or at a corner of the triangle; an array of outcoupling defects surrounding the triangle; and an external waveguide acting as a directional outcoupler. By exploring different configurations of defects and coupled waveguides, we show that the various properties of the lasing modes can be explained by, and are consistent with, the topological valley edge states of the VPC. We show that in a comparable cavity based on a conventionally designed photonic crystal defect waveguide (Extended Data Fig. 8), the lasing modes behave very differently: they tend to be localized and exhibit highly irregular mode spacings.

Our design consists of a triangular lattice of quasi-hexagonal holes drilled through the active medium of a THz QCL wafer, as shown in Fig. 1a. The lattice resembles a previous theoretical proposal for a VPC¹⁰, but with the dielectric and air regions inverted to account for the transverse-magnetic (TM) polarization of QCLs^{1,2}. With hexagonal holes, the lattice would be inversion-symmetric, and its band structure would have Dirac points at the Brillouin zone corners (K and K'). By assigning unequal wall-length parameters d_1 and d_2 (Fig. 1a), the inversion symmetry is broken, and bandgaps open at K and K'. Assuming negligible coupling between the K and K' valleys, the two gaps are associated with opposite Chern numbers $\pm 1/2$, meaning that they are topologically inequivalent. The Chern numbers switch sign upon swapping d_1 and d_2 (that is, flipping the hole orientations)¹⁰. We characterize the photonic band structure using three-dimensional

(3D) finite-element simulations (see Methods). With the lattice period $a = 19.5 \mu\text{m}$, the bulk band-structure has a gap from 2.99 THz to 3.38 THz (Fig. 1b). For a straight boundary between domains of opposite hole orientations, the projected band diagram has a gap spanned by edge states with opposite group velocities in each valley (Fig. 1c and Extended Data Figs. 1–4). These states are topologically protected provided that inter-valley scattering is negligible; this limitation is due to the overall T symmetry of the VPC¹⁰, and similar limitations apply to other photonic topological edge states (at THz or other frequencies) that do not rely on magnetic materials³. Figure 1d shows simulation results in which a wave launched at mid-gap frequency crosses a 120° corner with negligible backscattering (a scanning electron microscope (SEM) image of such a corner is shown in Fig. 1e). Near the domain wall (dashed line in Fig. 1e), the electric fields are concentrated in the QCL medium, which is favourable for lasing.

We patterned the lattice onto a THz QCL wafer (see Methods), with a domain wall forming a triangular loop of side length $21a$ (Fig. 2a). By design, the QCL wafer's gain bandwidth (approximately 2.95–3.45 THz; see Methods and Extended Data Fig. 5) overlaps with the photonic bandgap. Electrical pumping is applied only to the nearest three lattice periods on each side of the domain wall, to avoid supplying gain to bulk modes and to achieve low total pump current⁷. The in-plane modes are vertically outcoupled by scattering through the air holes drilled into the QCL active region, and through the defects described below. Calculating the eigenmodes with realistic material losses in the unpumped portion of the QCL medium (see Methods), we find regularly spaced high-Q eigenmodes at frequencies matching the previously computed bandgap (Fig. 2b). The typical eigenmode field distribution shows uniform electric field intensities along the domain wall, even at the sharp corners (top of Fig. 2c). We quantified the extended nature of the computed eigenmodes by showing that they have significantly lower inverse participation ratios along the domain wall, indicating less mode localization, compared with the eigenmodes of a conventional photonic crystal cavity of similar shape and size (see Methods and Extended Data Fig. 6).

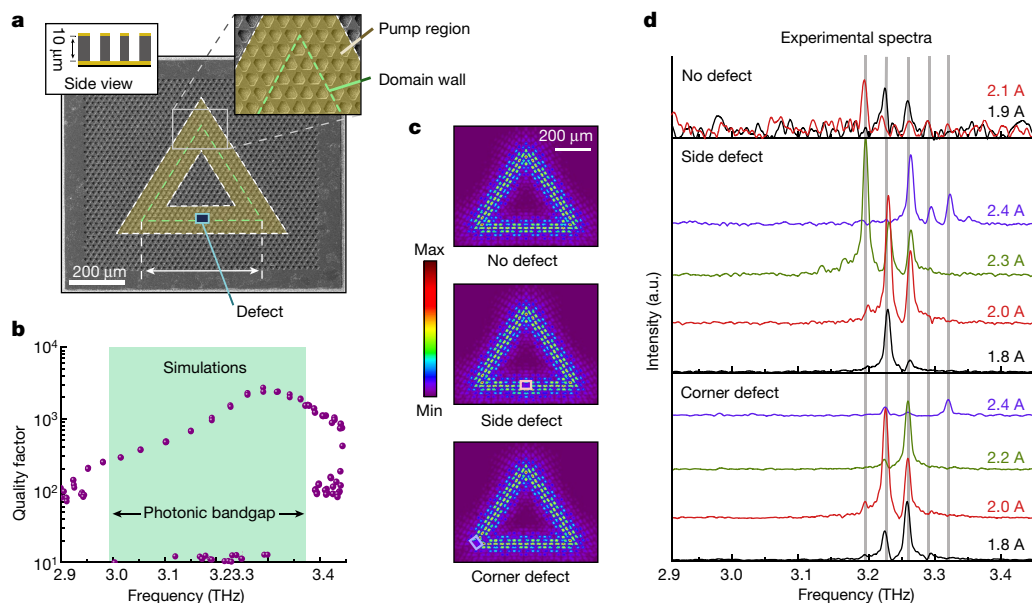


Fig. 2 | Fabrication and characterization of the topological THz QCL. a, SEM image of the THz QCL, whose optical cavity consists of an in-plane triangular loop of side length $21a$. The yellow shaded area is pumped by electrical injection, while the other parts are passive. The green dashed line indicates the domain wall. The black rectangle indicates a defect ($39 \mu\text{m} \times 33.5 \mu\text{m}$) etched right through the active medium of the THz QCL. Inset, cross-sectional schematic and magnified view of the domain wall. **b**, Calculated Q factors of the structure's eigenmodes, with realistic material absorption losses (approximately 20 cm^{-1}) within the passive region. The shaded area indicates

the regular spacing of the extended eigenmodes is a signature of running modes circulating around the triangular loop, analogous to whispering-gallery modes in a disk or a ring cavity³² (see Methods). This is the most striking feature imparted by the non-trivial topology of the VPC. The upper panel of Fig. 2d (labelled 'No defect') shows the experimentally measured emission spectra for this structure at two representative pump currents. There are regularly spaced peaks at 3.192 THz, 3.224 THz, 3.258 THz and 3.288 THz (vertical grey lines); the average free spectral range (FSR) is comparable to the FSR in the eigenmode simulations. The intensities are fairly low, owing to poor vertical outcoupling: the valley edge modes lie near K and K', below the light cone, so outcoupling occurs only by air-hole scattering. To improve the optical outcoupling efficiency (as well as to probe the robustness of the regular spacing against defects), we deliberately introduce a small rectangular defect, about $2a$ long and $\sqrt{3}a$ wide, drilled through the top metal plate and the active medium in the irregular cavity loop (Fig. 2a). Numerical simulations show that the defect has negligible effects on the field distributions (Fig. 2c) regardless of whether it is placed on an arm or a corner of the triangle. The resulting experimental lasing spectra exhibit substantially stronger peaks, with intensities enhanced by 10–20 times (see Extended Data Fig. 7, where the light-current-voltage characteristics of the topological lasers without an outcoupling defect, with a side defect, and with a corner defect show clearly the laser threshold and the 'roll-over' position of the QCL), while the emission peaks still maintain a regular spacing and have negligible frequency shifts relative to the original device (middle and bottom panels of Fig. 2d). The preservation of the peak frequencies indicates that the defect does not spoil the running-wave character of the lasing modes. With increasing pump current, we observe variations in the relative peak intensities. This 'mode-hopping' effect can be attributed to mode competition as well as to band structure realignment in the QCL wafer with the increase in the pump current; this is also observed in a conventional ridge laser fabricated on the same wafer (see Methods and Extended Data Fig. 5).

the photonic bandgap of the valley Hall lattice. **c**, Typical eigenmode electric field ($|E_z|$) profiles at around 3.23 THz, with no outcoupling defect, with a side defect, and with a corner defect. **d**, Emission spectra for the QCL with no outcoupling defect (top), with a side defect (middle), and with a corner defect (bottom). Grey vertical lines indicate the peak frequencies of the defect-free QCL, which correspond closely to those of the QCL with a defect. For clarity, the emission spectra are vertically offset with increasing pumping currents. a.u., arbitrary units.

For comparison, we fabricated a THz QCL with the same VPC design, but replaced the topological waveguide with a photonic crystal waveguide of size-graded holes, with all holes having the same orientation (Extended Data Fig. 8a). As before, a defect is introduced to improve the outcoupling efficiency. With a side defect on the arm of the triangular cavity, the experimental spectra exhibit multiple irregularly spaced lasing peaks between 3.20 THz and 3.38 THz (Extended Data Fig. 8d). When the defect position is moved to a corner of the triangular cavity, a completely new set of emission peaks is observed. Numerical simulations reveal numerous eigenmodes distributed over the upper half of the bandgap with a range of Q factors, no evident regular spacing patterns, and with modal intensities localized on different parts of the triangle (Extended Data Fig. 8c). This reflects the tendency of conventional waveguide modes to undergo localization, unlike the valley edge modes.

To probe the spatial distributions of the topological lasing modes and verify their running-wave nature, we fabricated another set of lasers that included an array of rectangular outcoupling defects arranged in a larger triangle enclosing the topological cavity (Fig. 3a). The defects are separated by a distance of several wavelengths (4λ) away from the domain wall and hence couple evanescently to the topological cavity lasing modes. We refer to the set of defects along each arm of the triangle as an 'emission channel'. By selectively blocking these emission channels (that is, covering the defects along certain arms), we can indirectly probe the spatial distributions of the lasing modes. When all emission channels are open, we observe regularly spaced emission peaks corresponding to topological lasing modes (Fig. 3b). Next, we sequentially cover two emission channels and measure the emission spectra from the remaining channel (Fig. 3a). In all three cases, the lasing spectra and the relative peak intensities under different pump currents are essentially the same (Fig. 3c–e), indicating that the lasing modes have equal intensities on the three arms of the triangular loop cavity.

The topological edge states form degenerate pairs circulating CW or CCW, which have the same intensity distributions, gain and vertical

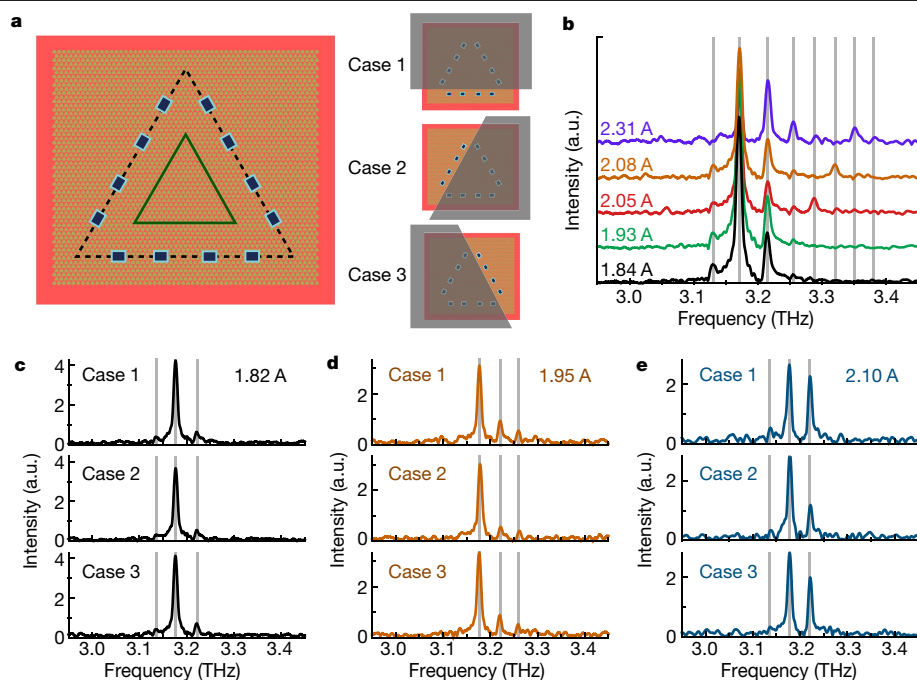


Fig. 3 | Topological laser with an array of evanescent outcouplers.

a, Schematic of the structure. A triangular loop cavity (green triangle) hosting topological edge states is surrounded by an array of outcoupling defects (blue rectangles) distributed around the perimeter of a larger triangle. The defects are eight lattice periods away from the topological interface, allowing for

evanescent outcoupling. The inset shows different defect-covering configurations for the spectral measurements. **b**, Emission spectra at different pump currents (vertically shifted for clarity), with all defects uncovered. **c–e**, Emission spectra at various pump currents for the three different defect-covering configurations shown in the inset of **a**.

outcoupling rates. Coupled-mode theory predicts that each topological lasing mode is composed of an equal-weight superposition of a CW and CCW pair (see Methods). The coexistence of CW and CCW modes

also explains why the defect along the cavity in Fig. 2 does not spoil the running-wave character, even in the presence of backscattering induced by the defect. To test this, we fabricated a sample with an additional

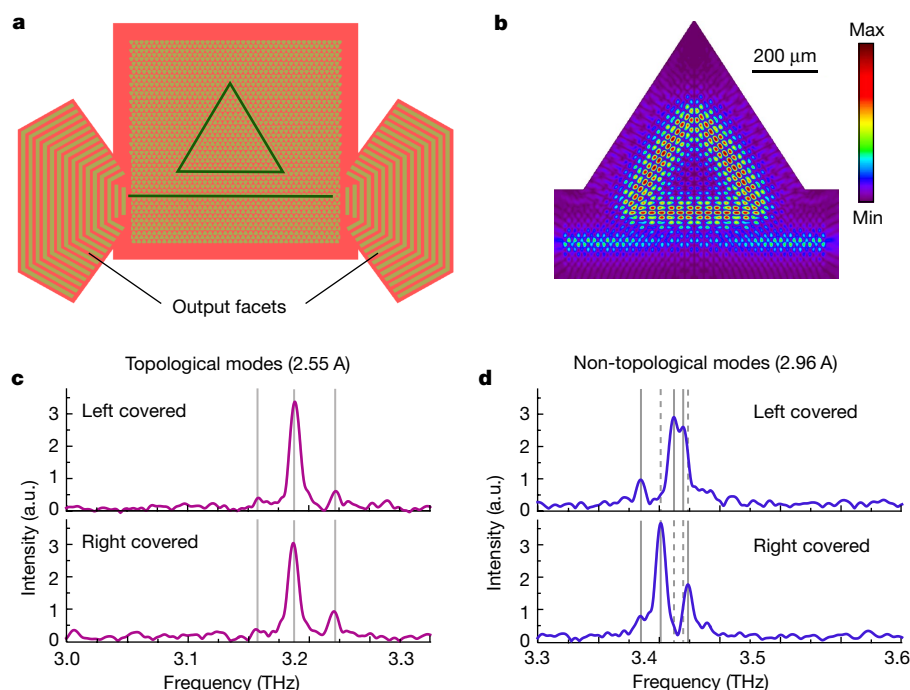


Fig. 4 | Topological laser in a directional outcoupling configuration.

a, Schematic of the structure. A straight valley edge-state waveguide is introduced below the bottom arm of the triangular loop cavity (topological interfaces are indicated by green lines), with outcoupling gratings on the left and right ends. The output facets are selectively covered to observe the directionality of the lasing modes. **b**, Intensity distribution for a typical

topological eigenmode obtained via a 3D numerical calculation. **c, d**, Emission spectra for the topological lasing modes (**c**) and non-topological lasing modes (**d**) with left and right output facets covered. For the topological lasing modes, the spectra have similar peak intensities, whereas for the non-topological lasing modes the spectra are completely different. The two sets of lasing peaks are measured under different pump currents.

straight topological waveguide located just below the triangular laser cavity (Fig. 4a). Each CW (CCW) cavity mode evanescently couples to the straight waveguide, propagates to the right (left) and then outcouples via a second-order grating. This sample is found to support three topological lasing modes with frequencies near 3.2 THz. By selectively covering the left or the right side of the device, we observe that each lasing mode emits with approximately equal intensities from the two facets (Fig. 4c and Extended Data Fig. 10a), indicating that the CW and CCW cavity modes have equal weights. For comparison, we observe that the same sample also supports non-topological lasing modes in a neighbouring frequency range, just above the photonic bandgap (around 3.4 THz), at high pumping currents, for example at 2.96 A. The non-topological lasing modes are observed to emit with very different intensities from the two output facets (Fig. 4d and Extended Data Fig. 10b). This demonstrates a qualitative difference in behaviour between topological and non-topological lasing modes in a single device.

In summary, we have implemented electrically pumped lasers based on the topological edge states of a valley photonic crystal, operating in the THz frequency regime. By investigating several different device configurations, we have established a chain of evidence demonstrating the running-wave features of the topological lasing modes. The most noteworthy observation is the regular mode spacing, which arises because the modes have running-wave characteristics despite the sharp corners of the cavity and various other disturbances. Looking ahead, there are further opportunities in using the valley degree of freedom in other active photonic devices, and the realization of an electrically pumped topological laser points the way towards incorporating topological protection into practical device applications. Apart from promising applications as a robust THz light source, this QCL platform may find immediate use in exploring the dynamical and nonlinear features of topological laser modes³³.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1981-x>.

1. Faist, J. et al. Quantum cascade laser. *Science* **264**, 553–556 (1994).
2. Köhler, R. et al. Terahertz semiconductor-heterostructure laser. *Nature* **417**, 156–159 (2002).
3. Hasan, M. Z. & Kane, C. L. Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
4. Ozawa, T. et al. Topological photonics. *Rev. Mod. Phys.* **91**, 015006 (2019).

5. Wittek, S. et al. Towards the experimental realization of the topological insulator laser. In *CLEO: QELS_Fundamental Science FTh1D-3* (Optical Society of America, 2017).
6. Bandres, M. A. et al. Topological insulator laser: experiments. *Science* **359**, eaar4005 (2018).
7. Harari, G. et al. Topological insulator laser: theory. *Science* **359**, eaar4003 (2018).
8. Bahari, B. et al. Nonreciprocal lasing in topological cavities of arbitrary geometries. *Science* **358**, 636–640 (2017).
9. Ju, L. et al. Topological valley transport at bilayer graphene domain walls. *Nature* **520**, 650–655 (2015).
10. Ma, T. & Shvets, G. All-Si valley-Hall photonic topological insulator. *New J. Phys.* **18**, 025012 (2016).
11. Gao, F. et al. Topologically protected refraction of robust kink states in valley photonic crystals. *Nat. Phys.* **14**, 140–144 (2018).
12. Schaibley, J. R. et al. Valleytronics in 2D materials. *Nat. Rev. Mater.* **1**, 16055 (2016).
13. Vitiello, M. S., Scarlari, G., Williams, B. & De Natale, P. Quantum cascade lasers: 20 years of challenges. *Opt. Express* **23**, 5167–5182 (2015).
14. Dhillon, S. S. et al. Terahertz transfer onto a telecom optical carrier. *Nat. Photonics* **1**, 411–415 (2007).
15. Gao, J. R. et al. Terahertz heterodyne receiver based on a quantum cascade laser and a superconducting bolometer. *Appl. Phys. Lett.* **86**, 244104 (2005).
16. Dean, P. et al. Terahertz imaging using quantum cascade lasers—a review of systems and applications. *J. Phys. D* **47**, 374008 (2014).
17. Sirtori, C., Barbieri, S. & Colombelli, R. Wave engineering with THz quantum cascade lasers. *Nat. Photonics* **7**, 691–701 (2013).
18. Zeng, Y., Qiang, B. & Wang, Q. J. Photonic engineering technology for the development of terahertz quantum cascade lasers. *Adv. Opt. Mater.* <https://doi.org/10.1002/adom.201900573> (2019).
19. Schröder, H. W., Stein, L., Frölich, D., Fugger, B. & Welling, H. A high-power single-mode CW dye ring laser. *Appl. Phys. (Berl.)* **14**, 377–380 (1977).
20. Pérez-Serrano, A., Javaloyes, J. & Balle, S. Longitudinal mode multistability in ring and Fabry–Pérot lasers: the effect of spatial hole burning. *Opt. Express* **19**, 3284 (2011).
21. Gordon, A. et al. Multimode regimes in quantum cascade lasers: from coherent instabilities to spatial hole burning. *Phys. Rev. A* **77**, 053804 (2008).
22. Hafezi, M., Demler, E. A., Lukin, M. D. & Taylor, J. M. Robust optical delay lines with topological protection. *Nat. Phys.* **7**, 907–912 (2011).
23. Peano, V., Houde, M., Marquardt, F. & Clerk, A. A. Topological quantum fluctuations and traveling wave amplifiers. *Phys. Rev. X* **6**, 041026 (2016).
24. Zhou, X., Wang, Y., Leykam, D. & Chong, Y. D. Optical isolation with nonlinear topological photonics. *New J. Phys.* **19**, 095002 (2017).
25. Barik, S. et al. A topological quantum optics interface. *Science* **359**, 666–668 (2018).
26. St-Jean, P. et al. Lasing in topological edge states of a one-dimensional lattice. *Nat. Photonics* **11**, 651–656 (2017).
27. Zhao, H. et al. Topological hybrid silicon microlasers. *Nat. Commun.* **9**, 981 (2018).
28. Dong, J. W., Chen, X. D., Zhu, H., Wang, Y. & Zhang, X. Valley photonic crystals for control of spin and topology. *Nat. Mater.* **16**, 298–302 (2017).
29. Kang, Y., Ni, X., Cheng, X., Khanikaev, A. B. & Genack, A. Z. Pseudo-spin–valley coupled edge states in a photonic topological insulator. *Nat. Commun.* **9**, 3029 (2018).
30. Shalaev, M. I., Walasik, W., Tsukernik, A., Xu, Y. & Litchinitser, N. M. Robust topologically protected transport in photonic crystals at telecommunication wavelengths. *Nat. Nanotechnol.* **14**, 31–34 (2019).
31. Lu, J. et al. Observation of topological valley transport of sound in sonic crystals. *Nat. Phys.* **13**, 369–374 (2017).
32. Sandoghdar, V. et al. Very low threshold whispering-gallery-mode microsphere laser. *Phys. Rev. A* **54**, R1777–R1780 (1996).
33. Seclí, M., Capone, M. & Carusotto, I. Theory of chiral edge state lasing in a two-dimensional topological system. *Phys. Rev. Res.* **1**, 033148 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Device fabrication, characterization and numerical simulations

We used THz QCL wafers with a three-well resonant-phonon GaAs/Al_{0.15}Ga_{0.85}As design, with the gain curve spanning 2.95 THz to 3.45 THz (ref.³⁴). The photonic crystal structures were patterned onto the wafer with a standard metal–semiconductor–metal configuration³⁵, as shown in Fig. 1a. The topological waveguide consists of quasi-hexagonal holes with opposite orientations on either side of the topological interface, with wall lengths $d_1 = 0.58a$ and $d_2 = 0.26a$ (or vice versa), where $a = 19.5 \mu\text{m}$ is the lattice period. The outcoupling defect for the sample shown in Fig. 2 consists of a rectangular hole with fixed size of $39 \mu\text{m} \times 33.5 \mu\text{m}$. The outcoupling defects for the sample shown in Fig. 3 consist of 12 rectangular holes of the same size, uniformly distributed along three triangle arms and situated eight lattice periods away from the topological interface.

The fabrication process began with metal (Ti/Au, 20/700 nm) deposition by an electron-beam evaporator onto the THz QCL wafer and an n⁺-doped GaAs host substrate, followed by Au/Au thermo-compression wafer bonding. Wafer polishing and selective wet etching using NH₃·H₂O/H₂O₂/H₂O (3/57/120 ml) solution were sequentially conducted to remove the THz QCL substrate down to an etch-stop layer. The etch-stop layer was then removed by 49% hydrofluoric acid solution, and the QCL active region was exposed for subsequent microfabrication. A 300-nm SiO₂ insulation layer was deposited onto the THz QCL wafer using plasma-enhanced chemical vapour deposition, followed by optical lithography and reactive-ion etching (RIE) to define the pumping area. The photonic structure patterns were transferred onto the THz QCL wafer by optical lithography, with deposition and lift-off of the top metal layer (Ti/Au, 20/900 nm). With the top metal layer as a hard mask, the photonic structures were formed by reactive-ion dry etching through the active region with a gas mixture of BCl₃/CH₄ = 100/20 standard cubic centimetres per minute. The top metal layer (remnant thickness approximately 300 nm) was retained as a top contact for current injection. The host substrate was covered by a Ti/Au (15/200 nm) layer as bottom contact. Finally, the device chip was cleaved, indium-soldered onto a copper heatsink, wire-bonded and attached to a cryostat cold finger for characterization.

The fabricated THz laser devices were characterized using a Bruker Vertex 70 Fourier-transform infrared spectrometer with a room-temperature deuterated-triglycine sulfate detector. Mounted in a helium-gas-stream cryostat with temperature control at 9 K, the devices were driven by a pulser with repetition rate of 10 kHz and pulse width of 500 ns. The emission signal was captured by the detector in the vertical direction and Fourier-transformed into a spectrum, with the spectrometer scanner velocity of 1 kHz and spectrum resolution of 0.2 cm⁻¹. To measure the emission from different outcouplers, for example, the rectangular outcoupling defects or gratings, a thin metal sheet (approximately 100 μm) coated with an absorptive PMMA layer (approximately 100 μm) was used to cover the device emission surface partially. The absorption layer (single-pass absorption rate approximately 40%) was coated to reduce the light reflection from the metal sheet. The cover was positioned using a custom stage with a positional accuracy of about 20 μm. The cover was placed very close to the device surface: the gap between the device surface and the metal sheet was smaller than 300 μm.

In this work, all numerical results were calculated using the finite-element method simulation software COMSOL Multiphysics. In 3D band diagram calculations, the 10-μm-thick QCL medium was modelled as a lossless dielectric with a refractive index of 3.6, sandwiched between metal layers modelled as perfect electrical conductors. All band structures were computed for TM polarization. The projected band diagram in Fig. 1c was obtained with a supercell with 10 quasi-hexagonal holes on each side of the domain wall; spurious modes localized at the boundaries of the computational cell were removed before plotting. In 3D

eigenmode calculations, the unpumped portion of the QCL medium was modelled as a lossy dielectric, accounting for the intrinsic loss of the actual semiconductor medium; the imaginary part of the refractive index is 0.0159, corresponding to an absorption loss of about 20 cm⁻¹. To reduce computational workload, eigenmodes were computed for a slightly smaller structure with several outermost unit cells removed, but with the triangular loop cavity left unchanged.

Valley photonic crystal design

Extended Data Fig. 1a shows the 2D band structure of a triangular-lattice photonic crystal whose unit cell comprises a regular hexagonal air holes in the dielectric of refractive index 3.6. This dielectric medium represents the QCL wafer medium in the actual device. The band structure exhibits Dirac points—linear band-crossing points between the two lowest TM photonic bands—at the corners of the hexagonal Brillouin zone, denoted by K and K'. Near K (K'), the Bloch states can be described by an effective 2D Dirac Hamiltonian^{36,37}:

$$H = v_D(\pm q_x \sigma_x + q_y \sigma_y) \quad (1)$$

where $\mathbf{q} = (q_x, q_y)$ is the wavevector measured from K (K'), v_D is the group velocity, $\sigma_{x,y}$ are the first two Pauli matrices, and the + (−) sign corresponds to K (K').

Setting $d_1 \neq d_2$ breaks the C_{3v} symmetry of the photonic crystal, and lifts the degeneracy of the Dirac points, as shown in Extended Data Fig. 1b. In Extended Data Fig. 1c, d, we plot the absolute values of the out-of-plane electric field $|E_z|$ and Poynting vectors within each unit cell at the K and K' points for both the lower band and upper band. The modes in the two valleys are time-reversed counterparts, as shown by the opposite circulations of electromagnetic power.

The effect of the symmetry breaking can be modelled as a mass term added in the effective Dirac Hamiltonian:

$$H = v_D(\pm q_x \sigma_x + q_y \sigma_y) + v_D m \sigma_z \quad (2)$$

where m represents the effective mass of Dirac particles, and σ_z is the third Pauli matrix. The band structures near the two valleys (that is, K and K') have identical dispersion but are topologically distinct. This can be shown by computing the valley-projected Chern number³, defined as

$$C_{K/K'} = \frac{1}{2} \int_{\text{HBZ}} \Omega_{K/K'}(\mathbf{q}) d\mathbf{S} \quad (3)$$

where the integration is performed only for half of the Brillouin zone (HBZ) containing K or K'. Here $\Omega_{K/K'}(\mathbf{q})$ is the Berry curvature defined as $\Omega = \nabla_k \times \mathbf{A}(k)$, where $\nabla_k = \left(\frac{\partial}{\partial k_x}, \frac{\partial}{\partial k_y} \right)$. $\mathbf{A}(k)$ represents the Berry connection, that is, $\mathbf{A}_n(k) = \int_{\text{unit cell}} d^2 r u_k^*(r) \nabla_r u_k(r)$, where $\nabla_r = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)$, and $u_k(r)$ represents the Bloch wavefunctions that can be calculated from numerical simulation.

Extended Data Fig. 2 shows the numerically calculated Berry curvature near K and K' points, whose integration over HBZ gives rise to opposite valley Chern numbers, that is, $C_K = 1/2$ and $C_{K'} = -1/2$. Rotating the quasi-hexagonal motif by 180° is equivalent to flipping the sign of the mass parameter m , which flips the signs of the valley Chern numbers ($C_{K'} = -1/2$, $C'_K = 1/2$).

Extended Data Fig. 3 shows a sample of photonic crystal consisting of two domains with opposite valley Chern numbers. The differences in valley Chern numbers between the two domains are

$$\Delta C_K = C_K - C'_K = -1; \quad \Delta C_{K'} = C_{K'} - C'_K = +1 \quad (4)$$

Thus, based on the topological bulk-boundary correspondence principle³, there shall be one forward-propagating edge state at K'

and one backward-propagating edge state at K. This is verified by the numerically calculated photonic band structure shown in Extended Data Fig. 3b. The field plots in Extended Data Fig. 3c, d show that the edge states are indeed strongly localized to the domain wall, that is, between the two domains with opposite valley Chern numbers.

Comparison of 2D and 3D band structures

In a 2D VPC with parameters stated in the main text, the bulk TM band structure has a bandgap from 3.23 THz to 3.51 THz (the relative bandwidth of around 8%), as shown by the black curves in Extended Data Fig. 4a. For a 2D structure with two domains of opposite hole orientations separated by a straight domain wall (such as in Extended Data Fig. 3a), the projected bandgap occupies a similar frequency range, and the valley edge states traverse the whole projected bandgap as shown by the black curves in Extended Data Fig. 4b.

In the actual experiment, the VPC is 3D, patterned onto a THz QCL wafer in a metal–semiconductor–metal configuration³⁵. The active medium is 10 μm thick, sandwiched between two metal plates to ensure subwavelength vertical confinement of the TM-polarized lasing waves within the active layer. Numerical results for the 3D structure are shown by the red curves in Extended Data Fig. 4. The band structure and projected band diagram are shifted to lower frequencies, but otherwise remain qualitatively similar.

Emission characteristics of conventional lasers (ridge laser and VPC laser)

To characterize the gain spectral range and other properties of the THz QCL wafer, we fabricated and studied a conventional ridge laser. Extended Data Fig. 5a plots the emission spectra at different pump currents. On scanning through the entire dynamic range of the pump, we observe that the gain spectral range is approximately 2.95 THz to 3.45 THz. With increasing pump, the emission spectrum envelope gradually blueshifts, which is due to the Stark shift of the intersubband transition in the THz quantum cascade medium^{38,39}.

To align the frequency of the VPC bandgap to the gain peak of the THz QCL (approximately 2.9–3.45 THz, evidenced by the range of emission peaks of the ridge laser), we fabricated a series of VPCs of various periods without any domain wall loop cavity. By studying the lasing peaks, we determined that the photonic bandgap of a VPC laser with $a = 19.50 \mu\text{m}$ and size of approximately $820 \mu\text{m} \times 725 \mu\text{m}$ extends from 2.99 THz to 3.39 THz, which is a good match for the gain peak range of the THz QCL wafer. These results also helped us to estimate the effective refractive index of the QCL active region to be around 3.60 at the operation frequency.

Extended nature of topological modes

The key feature of the topological laser cavity is that it supports whispering-gallery-like running-wave modes even in presence of the three sharp corners. By contrast, a trivial cavity cannot support such modes due to strong back-reflection at the corners, which localizes the electromagnetic field at various portions of the cavity.

This phenomenon can be quantified by calculating the inverse participation ratio (IPR) along the one-dimensional (1D) curve corresponding to the triangular loop. The IPR is widely used to characterize the localization of modes and is defined as⁴⁰

$$\text{IPR}(\omega) = \frac{\int_L |E_z(\omega, \xi)|^4 d\xi}{\left[\int_L |E_z(\omega, \xi)|^2 d\xi \right]^2} L \quad (5)$$

where ξ is the coordinate parametrizing the 1D curve of length L . The denominator in equation (5) ensures normalization. For a mode confined to a length L_0 , IPR goes as L/L_0 , whereas for completely delocalized modes $L_0 \approx L$, leads to $\text{IPR} \approx 1$; with increasing localization, L_0 decreases and therefore the IPR increases.

The numerical IPR results for the triangular loop cavity are shown in Extended Data Fig. 6. As expected, the topological modes have substantially smaller IPR than the non-topological modes.

Topological modes in the triangular loop cavity

Figure 2b of the main text shows the numerically calculated modes of a triangular cavity formed between two topologically inequivalent VPC domains. These high- Q modes are constructed out of topological edge states that have the characteristics of running waves.

From the condition that running waves should interfere constructively over each round trip, we can estimate the mode separation or the FSR. Constructive interference requires

$$\Delta k = \frac{2\pi}{L} \quad (6)$$

where k denotes the wavenumber for the running-wave-like envelope function corresponding to any given edge state, and L is the total path length (the circumference of the triangular loop). The edge states have an approximately linear dispersion relation $\Delta\omega = v\Delta k$, where ω is the angular frequency detuning relative to mid-gap and v is the group velocity. Hence, the FSR is

$$\Delta f = \frac{v}{L} \quad (7)$$

For the structure, $L \approx 1,257 \mu\text{m}$, and we estimate $v = 4.53 \times 10^7 \text{ m s}^{-1}$ from numerical calculations (Fig. 1c). This yields $\Delta f \approx 0.036 \text{ THz}$, which matches well with the simulations and the experimental results (for example, $\Delta f \approx 0.035 \text{ THz}$ for the simulation results shown in Fig. 2b, and $\Delta f \approx 0.033 \text{ THz}$ in the experimental results shown in Fig. 2d).

Owing to time-reversal symmetry, each running-wave mode has a degenerate counterpart with opposite circulation direction. Hence, modes can be constructed from superpositions of CW and CCW running waves. Numerical solvers typically do not return the CW and CCW solutions, but rather the superpositions of the two running waves. However, CW and CCW modes can be reconstructed from suitable superpositions of the degenerate solutions returned by the numerical solver (Extended Data Fig. 9).

The CW and CCW valley edge modes form two orthogonal basis modes and thus each topological lasing mode is a superposition of CW and CCW valley edge modes⁴¹. To determine the superpositions, we can use the framework of coupled-mode theory⁴². There are two important effects acting on the CW and CCW modes: weak coupling between CW and CCW modes, induced for example by symmetry-breaking defects in the VPC; and gain and loss, which are due to amplification by the gain medium, material dissipation and radiative outcoupling.

Using coupled-mode theory, we represent the states of the laser by $\psi = (ab)^T$, where a and b are the CW and CCW mode amplitudes respectively. The condition for steady-state lasing is

$$H_0\psi + i\left(\frac{g}{1+|\psi|^2} - \gamma\right)\psi = \delta\omega\psi \quad (8)$$

where

$$H_0 = \begin{bmatrix} 0 & -\kappa \\ -\kappa & 0 \end{bmatrix}$$

is a Hermitian Hamiltonian containing a coupling rate κ between the CW and CCW modes, both of which have zero frequency detuning, $\delta\omega$ is the frequency detuning of the steady-state lasing mode, g is the amplification rate due to the gain medium, and γ is the loss rate due to material dissipation and radiative outcoupling. Note that the gain is saturable.

Importantly, the non-Hermitian terms are diagonal because the CW and CCW modes are topologically protected running waves that have the same intensity distribution, and therefore should experience the same rates of gain and loss.

Regardless of the non-Hermitian terms, the solutions to the coupled-mode equation are

$$\begin{aligned}\psi &\propto \frac{1}{\sqrt{2}}(1 \ 1)^T \quad \text{for } \delta\omega = +\kappa \\ \psi &\propto \frac{1}{\sqrt{2}}(1 \ -1)^T \quad \text{for } \delta\omega = -\kappa\end{aligned}\quad (9)$$

In other words, the CW and CCW modes should contribute equally to the steady-state lasing mode. The overall amplitude can be determined by setting the imaginary part of the eigenproblem to zero.

These results hold not only at the lasing threshold, but also in the above-threshold regime where gain saturation is in effect. Above threshold, provided κ is not too large, a single steady-state lasing mode is spontaneously chosen from one of the two possible solutions solved above, and the other solution is suppressed (that is, its amplitude is pinned to zero) by gain competition.

The above analysis rests on the idea that the underlying a and b modes are counter-propagating topological modes. It does not apply if the modes experience different gain/loss rates (so that the non-Hermitian term is non-diagonal), or if they are non-degenerate—as is the case in the non-topological cavity, which lacks running-wave-like edge states.

Bidirectional outcoupling of laser modes

Here, we provide more details about the topological laser in the directional coupling configuration (Fig. 4 of the main text and Extended Data Fig. 10).

This structure features a straight topological waveguide placed below the triangular cavity (Fig. 4a). The valley Chern number difference along the straight waveguide is opposite to that along the bottom arm of the triangular cavity. Owing to valley conservation, a CW (CCW) cavity mode evanescently couples to a right- (left-) moving valley edge mode on the straight waveguide. The output facets on the left and right ends of the straight waveguide are second-order gratings. After using numerical simulations to optimize the grating parameters, the reflection ratio is estimated to be <10%, ensuring negligible light feedback into the straight waveguide and laser cavity.

Numerical simulations of the structure reveal topological eigenmodes at frequencies near 3.2 THz, within the topological gap of the VPC. The intensity plot for a typical eigenmode is shown in Fig. 4b. These numerically calculated topological eigenmodes are all twofold degenerate, consistent with the degenerate CW and CCW cavity modes of the triangular loop. Moreover, the structure hosts non-topological lasing modes around 3.4 THz, around the edge of the upper band. The non-topological modes are all non-degenerate.

In the experiment, each topological mode exhibits a ‘peak ratio’ (the ratio of emission peak intensities from two output facets) close to unity. A typical spectrum is shown in Fig. 4c, and the light-current curves are shown in Extended Data Fig. 10a. For the non-topological

modes, the peak ratios are far from unity (Fig. 4d and Extended Data Fig. 10b); for some of these, the peak is only clearly observable when one facet is covered but lies within the noise floor when the other facet is covered.

During repeated experimental runs with the same sample, we observe a repeatable set of peak frequencies for both the topological and non-topological lasing modes, but the exact peak intensities vary between runs due to the imprecise relative alignment of the covering metal sheet and sample. We observe that the topological modes have peak ratios close to unity, whereas the non-topological modes have different peak ratios.

Data availability

The data sets generated during and/or analysed during the current study are available in the DR-NTU(Data) repository <https://doi.org/10.21979/N9/PECAGQ>.

34. Belkin, M. et al. High-temperature operation of terahertz quantum cascade laser sources. *IEEE J. Sel. Top. Quantum Electron.* **15**, 952–967 (2009).
35. Williams, B. S., Kumar, S., Callebaut, H., Hu, Q. & Reno, J. L. Terahertz quantum-cascade laser at $\lambda \approx 100 \mu\text{m}$ using metal waveguide for mode confinement. *Appl. Phys. Lett.* **83**, 2124–2126 (2003).
36. Gao, Z. et al. Valley surface-wave photonic crystal and its bulk/edge transport. *Phys. Rev. B* **96**, 201402 (2017).
37. Wu, X. et al. Direct observation of valley-polarized topological edge states in designer surface plasmon crystals. *Nat. Commun.* **8**, 1304 (2017).
38. Vitiello, M. S. & Tredicucci, A. Tunable emission in THz quantum cascade lasers. *IEEE Trans. Terahertz Sci. Technol.* **1**, 76–84 (2011).
39. Fatholouloumi, S. et al. Terahertz quantum cascade lasers operating up to $\sim 200 \text{ K}$ with optimized oscillator strength and improved injection tunneling. *Opt. Express* **20**, 3866–3876 (2012).
40. Rockstuhl, C. & Scharf, T. *Amorphous Nanophotonics* (Springer, 2013).
41. Spreew, R. J. C., Neelen, R. C., van Druten, N. J., Eiel, E. R. & Woerdman, J. P. Mode coupling in a He-Ne ring laser with backscattering. *Phys. Rev. A* **42**, 4315–4324 (1990).
42. Suh, W., Wang, Z., & Fan, S. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE J. Quantum Electron.* **40**, 1511–1518 (2004).

Acknowledgements This work is supported by funding from the Singapore Ministry of Education (MOE), grants MOE2016-T2-1-128 and MOE2016-T2-2-159, and the National Research Foundation Competitive Research Program (NRF-CRP18-2017-02). U.C., Y.C. and B. Zhang acknowledge support from the Singapore MOE Academic Research Fund Tier 2, grants MOE2015-T2-2-008 and MOE2018-T2-1-022 (S), and the Singapore MOE Academic Research Fund Tier 3 grant MOE2016-T3-1-006. L.L., A.G.D. and E.H.L. acknowledge the support of the EPSRC (UK) HyperTerahertz programme (EP/P021859/1), and the Royal Society and the Wolfson Foundation.

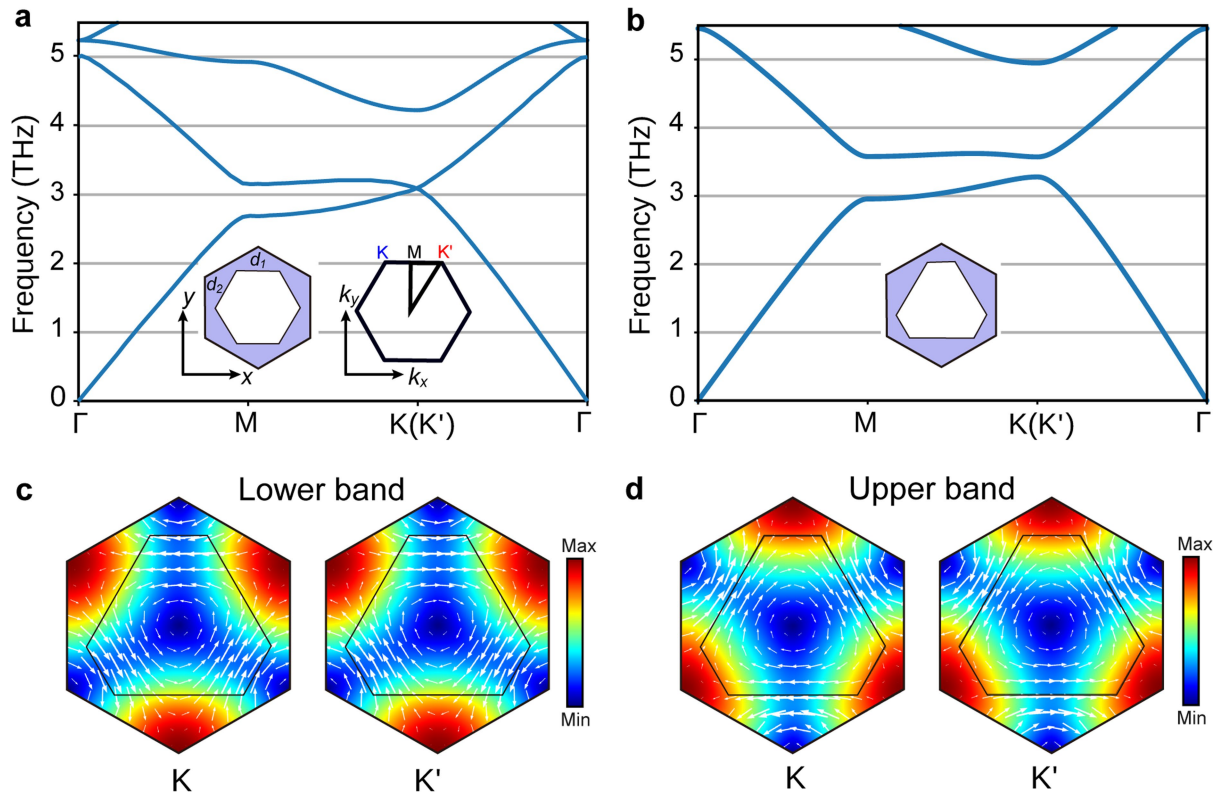
Author contributions Y.Z. and B.Q. fabricated the laser devices. Y.Z., J.L. and Y.J. performed the device characterization. L.L., A.G.D. and E.H.L. performed QCL wafer growth. Y.Z., U.C. and B. Zhu performed the simulations. Y.Z., U.C., B. Zhu, B. Zhang, Y.C. and Q.J.W. performed the theoretical analysis and contributed to manuscript preparation. B. Zhang, Y.C. and Q.J.W. supervised the project.

Competing interests The authors declare no competing interests.

Additional information

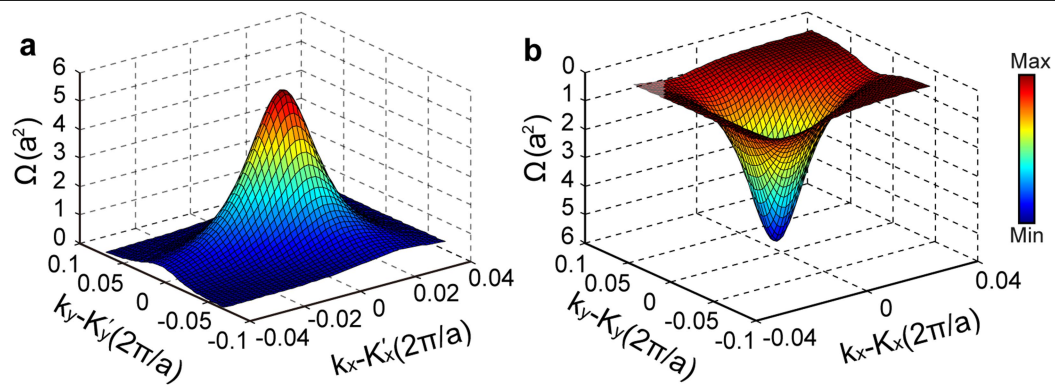
Correspondence and requests for materials should be addressed to B. Zhang, Y.C. or Q.J.W.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

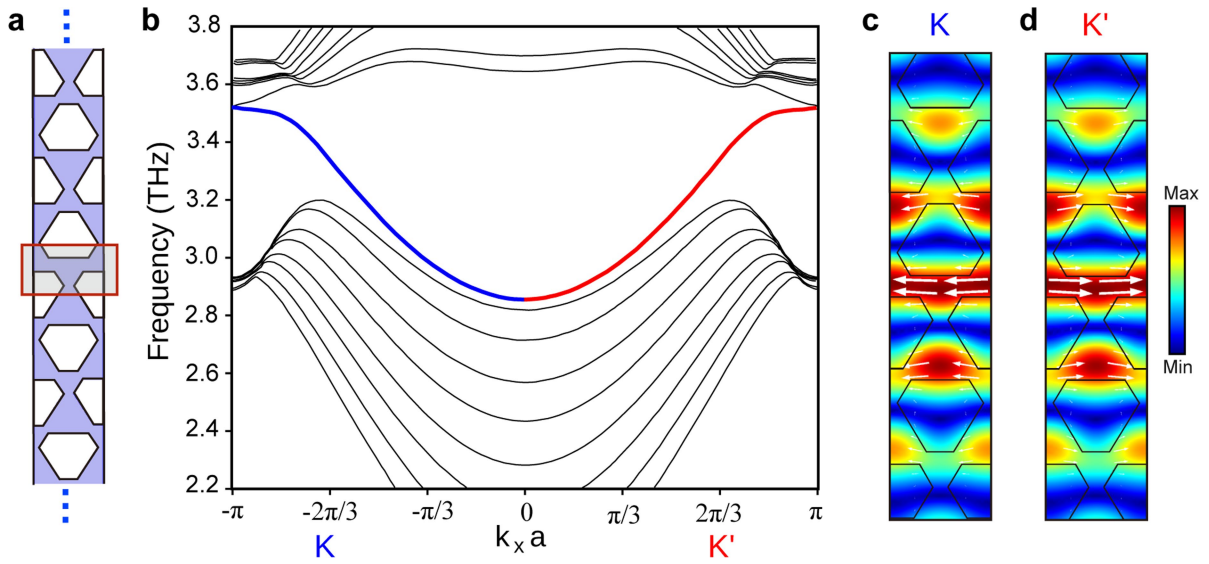


Extended Data Fig. 1 | Design of the 2D VPC. a, Photonic band structure for the TM modes of a 2D triangular photonic crystal of hexagonal air holes in dielectric (refractive index 3.6), with unbroken inversion symmetry. The unit cell and Brillouin zone are shown inset. **b**, Band structure after breaking inversion symmetry by setting $d_1 \neq d_2$. Inset, unit cell, with $d_1 = 0.58a$, $d_2 = 0.26a$.

The Dirac points at K and K' are lifted. **c, d**, Plots of the absolute value of the out-of-plane electric field $|E_z|$ (colour maps) and Poynting vector (white arrows) within each unit cell at the K and K' points. For both the lower band (**c**) and upper band (**d**), the modes in the two valleys are time-reversed counterparts, as shown by the opposite circulations of electromagnetic power.

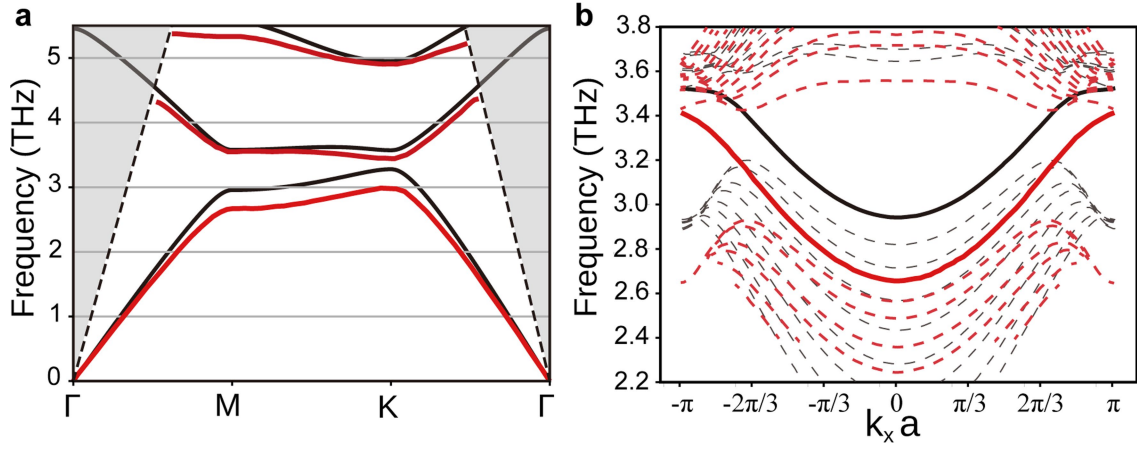


Extended Data Fig. 2 | Berry curvatures calculated using 2D Bloch wavefunctions for the lowest TM band. a, Near the K' valley. b, Near the K valley.



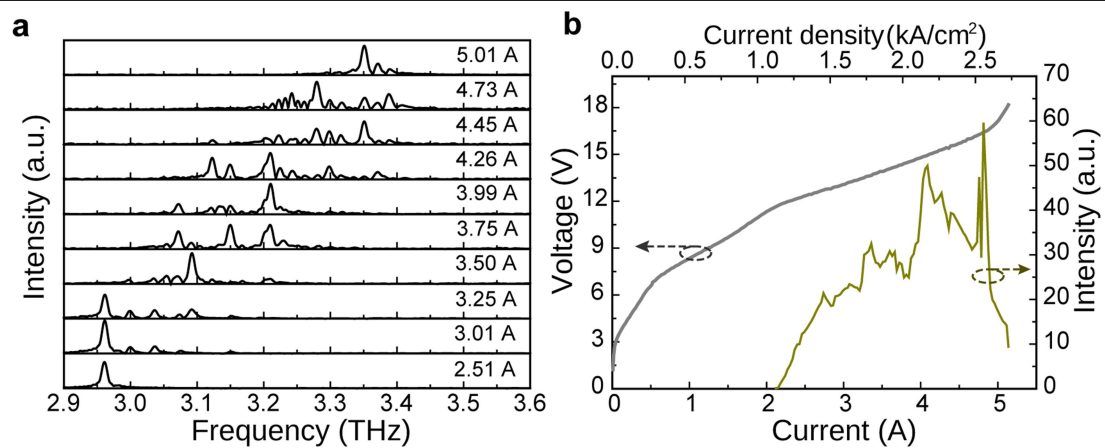
Extended Data Fig. 3 | Edge states of the 2D VPC. **a**, Supercell comprising two inequivalent VPC domains separated by a domain wall (highlighted by a red box). **b**, Projected band diagram for the supercell. The red (blue) curve

indicates the valley edge mode for the K (K') valley. **c**, **d**, Out-of-plane electric field $|E_z|$ (colour maps) and Poynting vector (white arrows) for the edge modes at K, K'.

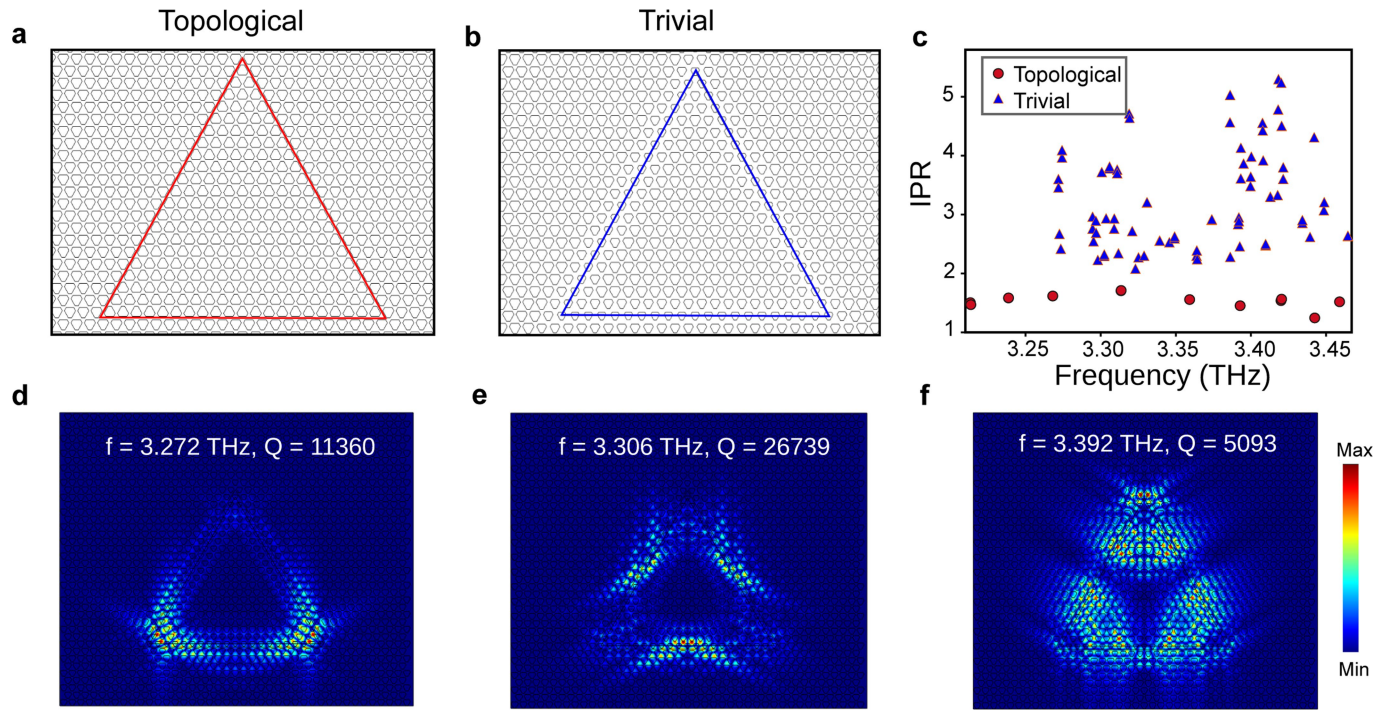


Extended Data Fig. 4 | Comparison between 2D and 3D TM photonic band structures. a, Bulk band structures of the 2D VPC (black) and 3D VPC (red). The grey regions delimited by black dashes denote the light cone. The 2D VPC is regarded as infinite in the out-of-plane (z) direction. The 3D VPC is modelled after the experiment, that is, metal-semiconductor-metal heterostructure

with central dielectric thickness of $10\ \mu\text{m}$. **b,** Projected band diagrams for a topological waveguide in 2D (black) and 3D (red). The lattice configuration is the same as in Extended Data Fig. 3a, with 10 quasi-hexagonal holes on each side of the domain wall. The edge states are plotted as thick solid curves for clarity.



Extended Data Fig. 5 | Emission characteristics of a conventional ridge laser fabricated on the quantum cascade wafer. a, Emission spectra at different pump currents. **b,** Light-current-voltage curves of the ridge laser.

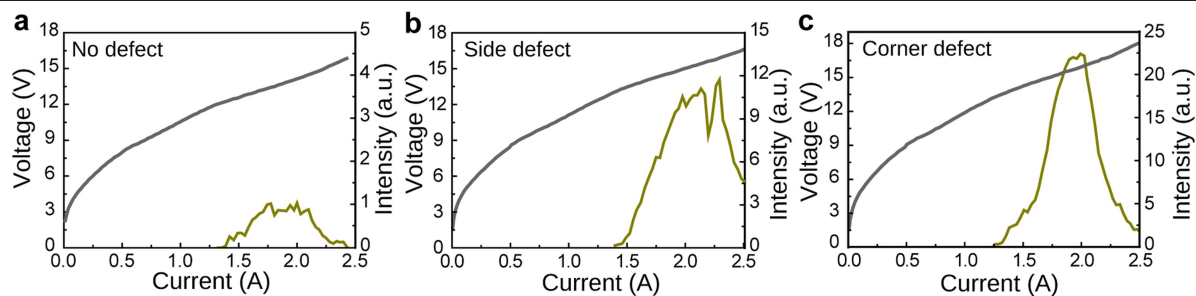


Extended Data Fig. 6 | IPR for trivial and topologically non-trivial modes.

a, b, Schematics showing the topologically non-trivial (**a**) and trivial (**b**) cavities. The 1D interfaces along which the IPR is calculated are indicated by red and blue lines. For the design of the trivial cavity, see Extended Data Fig. 8a.

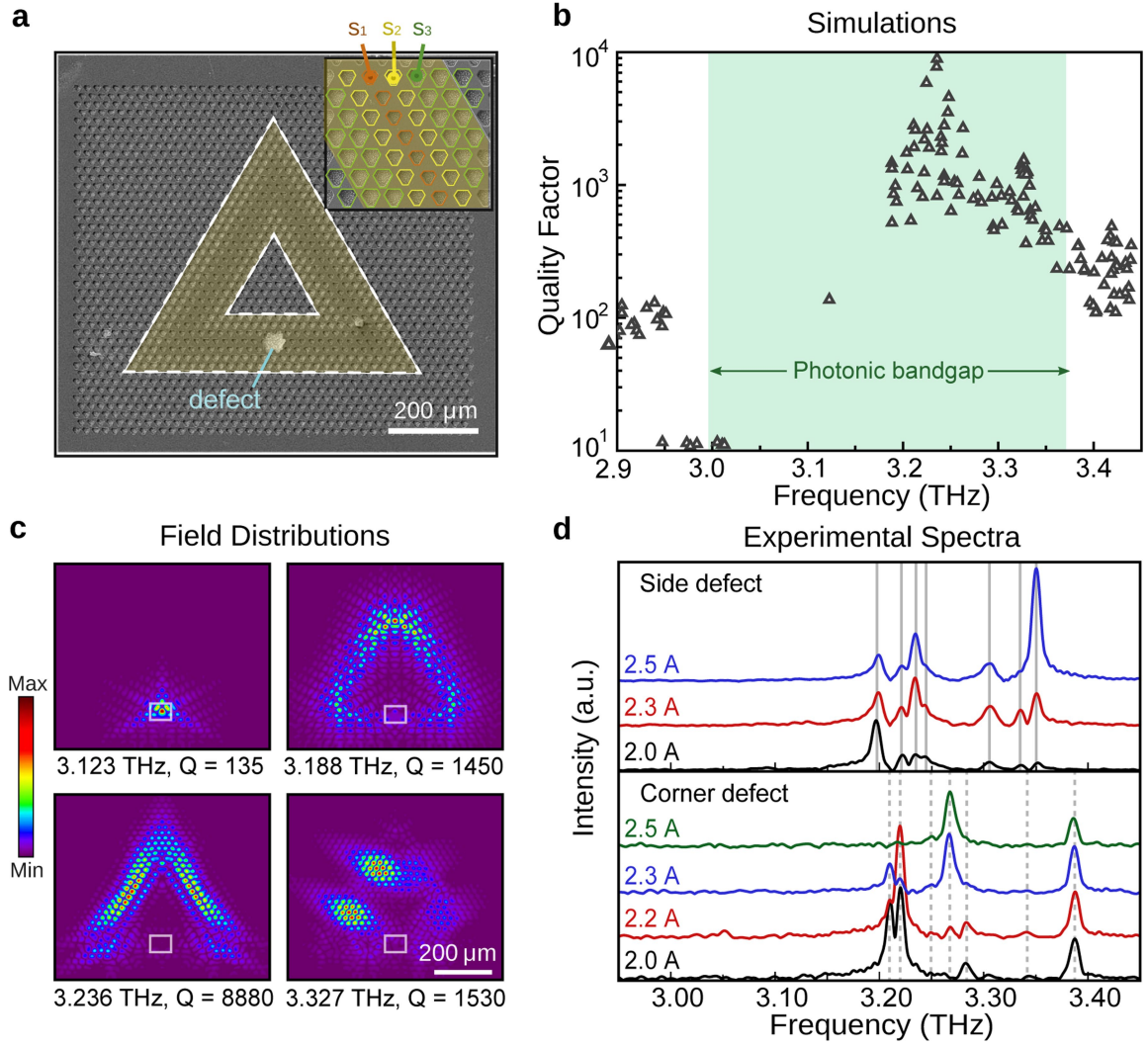
c, IPR versus frequency for eigenmodes in the band gap for each type of cavity.

The topological cavity's eigenmodes have consistently lower IPR, indicating that they are more uniformly extended along the loop. **d–f,** Intensity distributions for three representative eigenmodes of the trivial cavity. For comparison, eigenmodes of the topological cavity are shown in Fig. 2c (top) of the main text.



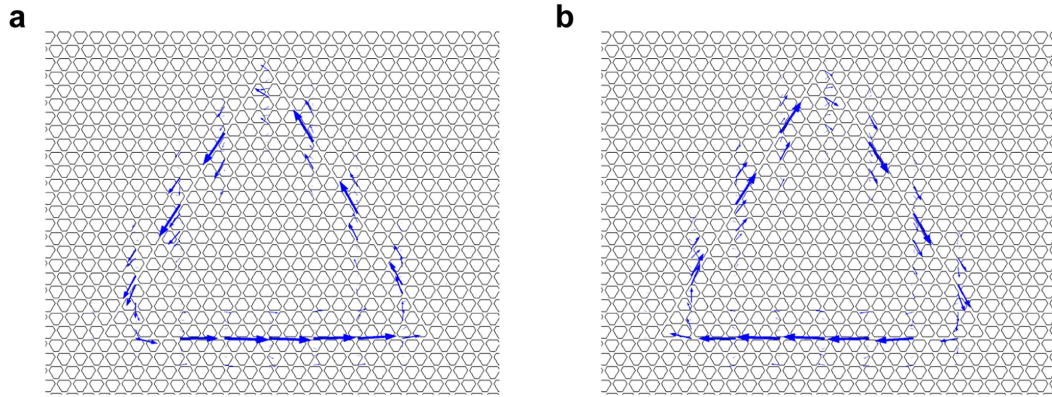
Extended Data Fig. 7 | Light-current-voltage curves of the topological laser with different designs. **a.** The topological laser without an outcoupling defect. **b.** The topological laser with a side defect. **c.** The topological laser device with a corner defect. The corresponding device emission spectra are shown in Fig. 2d.

All intensities in three sub-figures are measured with the same intensity scale. It can be inferred from these curves that the emission power is greatly enhanced by the outcoupling defect.



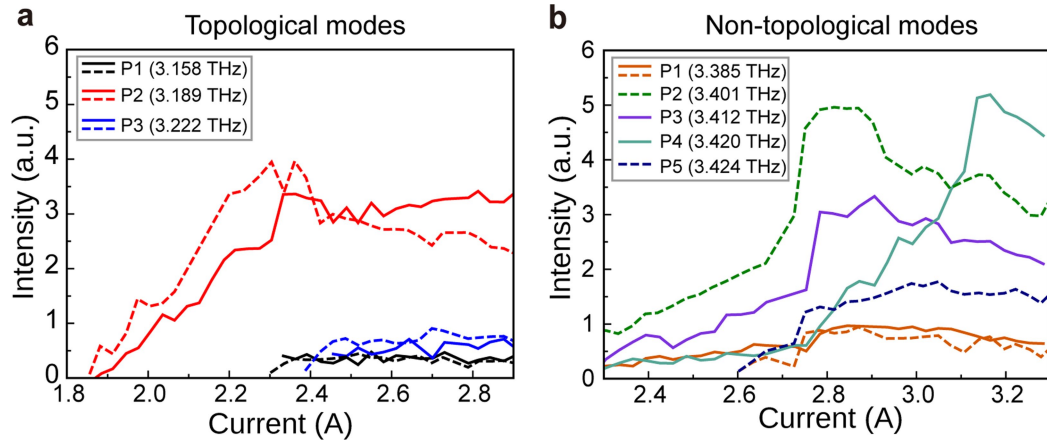
Extended Data Fig. 8 | Topologically trivial laser with triangular loop cavity formed by a conventional photonic crystal waveguide. **a**, SEM image of the fabricated structure. Inset, close-up view of the waveguide with single hole orientation, which consists of five rows of size-graded holes (with size scale factors $s_1 = 0.77$, $s_2 = 0.87$, $s_3 = 1$). A defect ($39 \mu\text{m} \times 33.5 \mu\text{m}$) is included to improve outcoupling efficiency. **b**, Calculated eigenmode Q factors for the structure with a side defect. The shaded area indicates the photonic bandgap

of the valley Hall lattice. **c**, Electric field ($|E_z|$) plots for typical calculated eigenmodes of the trivial cavity. The white square indicates the position of the side defect. **d**, Emission spectra of the topologically trivial lasers with a side defect (top panel) and corner defect (bottom panel) at different pump currents. The spectra are vertically offset for clarity. The emission peaks of two lasers are different and do not present a clear and regularly spaced pattern in frequency space.



Extended Data Fig. 9 | Quiver plots of Poynting vectors for two degenerate modes in a topologically non-trivial triangular loop cavity. a, b, Starting from two degenerate eigenmodes returned by the numerical solver, denoted

by ψ_1 and ψ_2 , the plotted modes are (a) $\psi_1 + i\psi_2$ and (b) $\psi_1 - i\psi_2$. These have CCW and CW characteristics, respectively.



Extended Data Fig. 10 | Lasing peak intensity curves for topological and non-topological lasing modes in the same laser device in a directional outcoupling configuration. The schematic of the device is shown in Fig. 4a of the main text. **a, b,** Here, peak intensities are plotted versus pump current for the topological modes (**a**) and non-topological modes (**b**) of the same sample. P1, P2 and so on represent different emission peaks. Solid (dashed) curves

correspond to the measurement with left (right) side of the device covered. Emission spectra at two representative pump currents are shown in Fig. 4c,d of the main text. For the topological lasing modes, the spectra from two output facets have comparable peak intensities, whereas for the non-topological lasing modes the peaks differ in intensity and frequency in the two cases.

Li metal deposition and stripping in a solid-state battery via Coble creep

<https://doi.org/10.1038/s41586-020-1972-y>

Received: 4 May 2018

Accepted: 1 November 2019

Published online: 3 February 2020

Yuming Chen^{1,2,3,10}, Ziqiang Wang^{1,2,10}, Xiaoyan Li^{1,2,3,4}, Xiahui Yao^{1,2}, Chao Wang^{1,2}, Yutao Li^{5,6}, Weijiang Xue^{1,2}, Daiwei Yu⁷, So Yeon Kim^{1,2}, Fei Yang^{1,2}, Akihiro Kushima⁸, Guoge Zhang⁴, Haitao Huang⁴, Nan Wu^{5,6}, Yiu-Wing Mai⁹, John B. Goodenough^{5,6} & Ju Li^{1,2*}

Solid-state lithium metal batteries require accommodation of electrochemically generated mechanical stress inside the lithium: this stress can be^{1,2} up to 1 gigapascal for an overpotential of 135 millivolts. Maintaining the mechanical and electrochemical stability of the solid structure despite physical contact with moving corrosive lithium metal is a demanding requirement. Using in situ transmission electron microscopy, we investigated the deposition and stripping of metallic lithium or sodium held within a large number of parallel hollow tubules made of a mixed ionic–electronic conductor (MIEC). Here we show that these alkali metals—as single crystals—can grow out of and retract inside the tubules via mainly diffusional Coble creep along the MIEC/metal phase boundary. Unlike solid electrolytes, many MIECs are electrochemically stable in contact with lithium (that is, there is a direct tie-line to metallic lithium on the equilibrium phase diagram), so this Coble creep mechanism can effectively relieve stress, maintain electronic and ionic contacts, eliminate solid–electrolyte interphase debris, and allow the reversible deposition/stripping of lithium across a distance of 10 micrometres for 100 cycles. A centimetre-wide full cell—consisting of approximately 10¹⁰ MIEC cylinders/solid electrolyte/LiFePO₄—shows a high capacity of about 164 milliampere hours per gram of LiFePO₄, and almost no degradation for over 50 cycles, starting with a 1× excess of Li. Modelling shows that the design is insensitive to MIEC material choice with channels about 100 nanometres wide and 10–100 micrometres deep. The behaviour of lithium metal within the MIEC channels suggests that the chemical and mechanical stability issues with the metal–electrolyte interface in solid-state lithium metal batteries can be overcome using this architecture.

Demands for safe, dense energy storage provide incentive for the development of all-solid-state rechargeable Li metal batteries^{3–5}. (Lithium metal batteries are to be distinguished from lithium ion batteries, in which the anode does not contain metallic lithium.) Lithium in the body-centred cubic (b.c.c.) crystal structure has 10× the gravimetric capacity and 3× the volumetric capacity of graphite⁶. The problem is that the non-lithium-metal volume fraction ϕ , consisting of entrapped solid–electrolyte interphase (SEI) debris, pores and other ancillary/host structures, tends to increase with battery cycling^{7–12}. Once a Li-metal-containing anode has $\phi > 70\%$, it loses its volumetric advantage compared to a graphite anode. Most solid electrolytes are thermodynamically unstable in contact with the corrosive Li metal¹³, forming SEI at a fresh solid electrolyte/Li metal interface. This thermodynamic instability can be predicted by checking the equilibrium phase diagram: it occurs when the solid electrolyte phase does not have a direct tie-line

connecting to the Li_{bcc} phase. Ab initio calculations have shown that a small number of compounds such as LiF, LiCl and Li₂O are absolutely stable against Li metal, but they are poor ionic conductors¹³. Good solid electrolytes (ionic conductors but electronic insulators) will decompose upon contact with Li_{bcc} to form SEI. Under large fluctuating mechanical stresses the SEI and the solid electrolyte can spall off and get entangled with Li: and as they are electronic insulators, they can cut off electronic percolation and cause ‘dead lithium’. The dual requirements of maintaining contact and adhesion with moving Li without fracture (mechanical stability) while reducing SEI production (electrochemical stability) makes the problem hard from an electrochemo-mechanics perspective.

Metallic lithium has a volume $\Omega = 21.6 \text{ \AA}^3$ per atom = 0.135 eV per GPa. This means that an overpotential U of -0.135 V , which is frequently seen experimentally in Li deposition, can in principle generate GPa-level

¹Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ³College of Environmental Science and Engineering, Fujian Normal University, Fuzhou, China. ⁴Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong, China. ⁵Texas Materials Institute, The University of Texas at Austin, Austin, TX, USA. ⁶Materials Science and Engineering Program, The University of Texas at Austin, Austin, TX, USA. ⁷Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁸Advanced Materials Processing and Analysis Center, Department of Materials Science and Engineering, University of Central Florida, Orlando, FL, USA. ⁹Centre for Advanced Materials Technology (CAMT), School of Aerospace, Mechanical and Mechatronics Engineering, The University of Sydney, Sydney, New South Wales, Australia. ¹⁰These authors contributed equally: Yuming Chen, Ziqiang Wang. *e-mail: liju@mit.edu

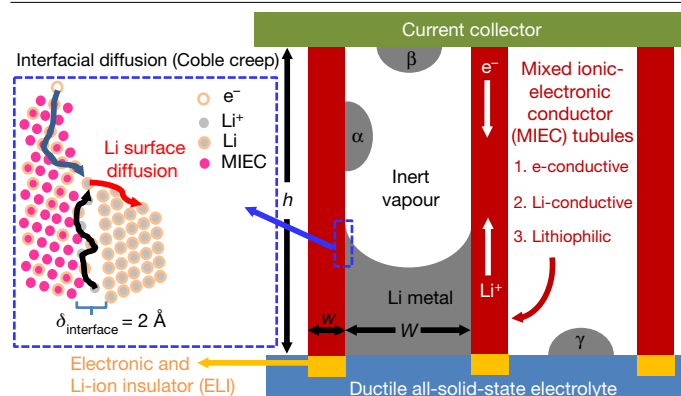


Fig. 1 | Mixed ionic-electronic conductor (MIEC) tubules as 3D Li hosts.

Schematic process of creep-enabled Li deposition/stripping in an MIEC tubular matrix with a geometry of $\{h, W, w\}$, where Coble creep dominates via interfacial diffusion along the MIEC/ Li_{bcc} incoherent interface. Main panel, cross-section of the matrix: MIEC tubules are shown as red, with white arrows indicating the free movements of electrons (e^-) and lithium ions (Li^+); the three required properties of the MIEC (red arrow) are labelled 1, 2 and 3. An electronic and Li-ion insulator (ELI; yellow) material is used at the root of the MIEC as a 'binder' to the solid electrolyte ('Ductile all-solid-state electrolyte'). α , β and γ are Li_{bcc} drops that are still recoverable. The boxed area is shown expanded in the inset: see Methods section 'Quantitative analysis' for details.

hydrostatic pressure (P_{LiMetal}) in Li according to the Nernst equation^{2,14}, and this stress will be transmitted to the surrounding solid structure. If these electrochemically generated mechanical stresses are not relieved, Li fingers or wedges may crack the solid electrolyte², through its grain boundaries or through its lattice. As the crack tip may be closer to the cathode, there is a transport advantage to the deposition of more Li at the crack tip, so $P_{\text{LiMetal}}(\mathbf{x})$ is generated again at the crack tip where \mathbf{x} is spatial position, and the process repeats until electrical shorting happens¹. The well-known elastic-modulus-based criterion⁵ for mechanical stability is not applicable when considering this crack-based degradation mode. The potentially huge electrochemically generated stress $P_{\text{LiMetal}}(\mathbf{x})$, if not relaxed quickly by creeping of Li, would fracture solid components. This and the chemical attack leading to SEI production makes the architecture of all-solid-state rechargeable Li metal batteries difficult to construct, even at a conceptual level.

Because Li melts at $T_{\text{M}} = 180^\circ\text{C}$ and is a soft metal, an alternative concept is to have the Li flow into and out of a 3D tubular structure like that shown in Fig. 1^{15–17}, keeping contact with a 3D solid host structure made of mixed ionic-electronic conductor (MIEC) that is absolutely electrochemically stable against Li metal (that is, having a direct tie-line to Li_{bcc} phase on the equilibrium phase diagram without intervening phases). Such 3D host structures have been studied experimentally in the past^{15–17}, but here we seek quantitative mechanistic understanding for the plating/stripping behaviour. We note that, in our construction, we choose only the MIEC (not the solid electrolyte) to be in thermodynamic equilibrium with Li, so it will not generate any SEI upon contact with Li, removing the possibility of SEI and SEI-based degradation. At 300 K, the homologous temperature for Li is $T/T_{\text{M}} = 0.66$, so Li should manifest an appreciable creep strain rate $\dot{\epsilon}(T, \sigma)$ (where σ is the deviatoric shear stress) by dislocation power-law creep or diffusional creep mechanisms, according to the deformation mechanism map of metals^{18,19}. Creep imparts an effective viscosity $\eta \equiv \sigma/\dot{\epsilon}(T, \sigma)$, so the Li may behave like an 'incompressible work fluid', and advancement and retraction of pure Li may be established inside the MIEC tubules (which cannot chemically react with the corrosive work fluid), driven by the chemical potential/pressure gradient $-\Omega \nabla P_{\text{LiMetal}}(\mathbf{x})$. The competition between interfacial-diffusional Coble creep, bulk diffusional Nabarro–Herring creep, and hybrid diffusive-displacive dislocation

creep mechanisms depends on the grain size. The pore space helps to relieve the stresses (hydrostatic and deviatoric) by allowing the Li_{bcc} to backfill by diffusion, so the all-solid-state host is not fractured during cycling (ensuring mechanical stability), while maintaining high-quality electronic and ionic contacts. While few solid electrolytes with satisfactory Li-ion conductivity are electrochemically stable against Li_{bcc} (ref. 13), there exist many MIECs with such stability and they will not decompose to form fresh SEI at the MIEC/metal interface. These include popular anode materials like lithiated graphite or hard carbon (LiC_6 is an MIEC), Si ($\text{Li}_{22}\text{Si}_5$ is an MIEC), Al (Li_9Al_4 is an MIEC) and so on²⁰, as well as materials with appreciable solubility of Li atoms as a random solid solution (CuLi_x) or even bulk-immiscible metals (such as $\text{M} = \text{Ni}, \text{W}$) that may nonetheless support some Li solubility at the M/Li_{bcc} phase boundary. Here we focus on lithiated carbonaceous materials as the MIEC 'rail' that guides Li_{bcc} deposition and stripping, although in the 'Quantitative analysis' section of Methods we show that this design is almost independent of MIEC material when using channels about 100 nm wide and 10–100 μm deep.

The cycling of Li under alternating negative and positive overpotential is rather like the application of a pump, which can produce fatigue in the solid host structure. To avoid such fatigue, the MIEC walls should be sufficiently strong and ductile to accommodate the stresses generated by P_{LiMetal} and capillarity. Typical graphene foam with too thin a wall thickness w may not be appropriate because such walls may easily tear, crumble or fold due to van der Waals adhesion. Also, the contact condition between the MIEC and the solid electrolyte capping layer is important, as this is where Li deposition is most likely to occur initially and where P_{LiMetal} is initiated. A root or coating of an electronic and Li-ion insulator (ELI) material like BeO , SrF_2 or AlN (with a bandgap $>4.0\text{ eV}$, and thermodynamically stable against Li_{bcc}) might be used to bind the MIEC to the solid electrolyte. A mechanically compliant solid electrolyte, for example polyethylene oxide (PEO), could be used to prevent the brittle root-fracture problem.

In the following experiments, we use lithiated carbon tubules $\sim 100\text{ nm}$ wide as the MIEC material. We demonstrate plating/stripping of Li or Na inside individual carbon tubules in an in situ transmission electron microscope (TEM) experiment, where a PEO-based polymer about 50 μm thick was used as the solid electrolyte. The opposite side of the solid electrolyte was coated with a Li counter-electrode connected to the scanning tunnelling microscope (STM)/TEM manipulator. The TEM copper grid (Fig. 2a and Supplementary Fig. 1) serves as the current collector attached to the carbon tubules on the other end. The carbon tubule has an inner diameter W of around 100 nm, and its walls of width $w \approx 20\text{ nm}$ are also nanoporous, as shown in Fig. 2b and Supplementary Fig. 2a, b²¹.

Figure 2b–d shows TEM images of the Li plating process in a single carbon tubule with ZnO_x as a lithiophilic agent introduced by controlling the synthetic process (see Methods and Supplementary Video 1). Figure 2e, f and Supplementary Video 2 show the changes of selected-area electron diffraction (SAED) patterns when deposited Li passes through the original void. After the ring pattern of the carbon tubule and a period of changing SAED patterns, the SAED (Fig. 2f) stays stable and shows a strong texture: $(110)_{\text{Li}_{\text{bcc}}} \perp$ tubule axis and $(\bar{1}\bar{1}0)_{\text{Li}_{\text{bcc}}} \parallel$ tubule axis (Supplementary Fig. 3 also demonstrates the single-crystal feature). Moreover, a high-resolution TEM (HRTEM) video captures the first appearance of the fresh Li crystal, with a 0.248-nm lattice spacing measured between (110) crystal planes perpendicular to the wall (Fig. 2g–i and Supplementary Video 3). We decreased the electron beam current to 0.3 A cm^{-2} to maintain the HRTEM image of the Li crystal for several seconds. The Li can also be stripped along the tubule (Supplementary Fig. 4) by retracting the Li_{bcc} tip. The tip can plate and strip a length of more than 6 μm along the carbon tubule, which was the largest unblocked length of carbon tubule we could find (Supplementary Figs. 5 and 6). It can even climb over partial obstructions inside the carbon tubule (Supplementary Fig. 7). We also discovered

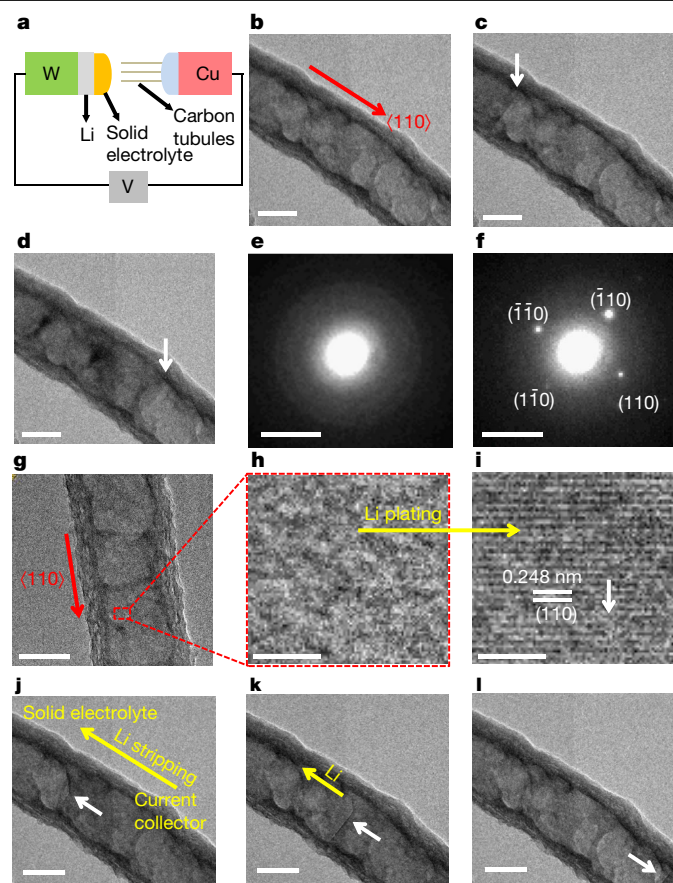


Fig. 2 | Lithium plating/stripping inside carbon tubules. **a**, In situ TEM set-up: see Methods for details. **b–d**, TEM imaging of Li plating with fronts marked by white arrows (Supplementary Video 1) at increasing time. **e, f**, SAED changes from **e** to **f** during Li plating (Supplementary Video 2). **g–i**, HRTEM imaging of a tubule before plating (**g**; boxed region shown magnified in **h**) and after plating (**i**, showing first formation of a Li crystal (Supplementary Video 3)). The red arrow indicates the Li atomic transport direction in deposition, pointing from the solid-electrolyte to the current collector. **j–l**, TEM imaging of Li stripping with a void plug between Li and solid electrolyte, with the Li atomic transport direction indicated by the yellow arrow (from current-collector side to the solid-electrolyte side), and the surface extent indicated by white arrows (Supplementary Video 4). Scale bars: **b–d, g, j–l**, 100 nm; **h, i**, 2 nm; and **e, f**, 5 nm^{−1}.

reversible Li plating/stripping in aligned double carbon tubules (Supplementary Fig. 8 and Supplementary Video 5). The lithium filling ratio inside the tubule was estimated by electron energy-loss spectroscopy (EELS) thickness measurement. The Li K-edge of EELS (Supplementary Fig. 8c, d) is observed after Li plating at the location shown by a red cross in Supplementary Fig. 8a^{7,22}. The diameter of the plated Li_{bcc} plug is estimated to be 92 nm, which can be compared to the inner diameter of the tubule, about 100 nm. Li_{bcc} can also be plated in three aligned tubules simultaneously (Supplementary Fig. 9).

We have tested the cycling stability of the carbonaceous MIEC tubules by in situ TEM, and found they can maintain excellent structural integrity even after 100 cycles of Li plating and stripping (Supplementary Video 6, Supplementary Figs. 10, 11). Li can also plate/strip inside tubules of different sizes, and even within tubules filled with 3D obstacles (Supplementary Figs. 7, 12–14). Our observations indicate that the internal Li shape change is not displacive/convective, but rather a diffusive plating/stripping process onto a front or fronts, which is much more tolerant of internal obstructions or obstacles. Similar results for plating/stripping sodium metal are shown in Supplementary Figs. 15, 16 and Supplementary Videos 7, 8.

When stripping Li (Supplementary Video 4, Fig. 2j–l and Supplementary Fig. 17), we can sometimes create a void plug that grows between the residual Li and the solid electrolyte. Yet this gap does not prevent the Li from being further stripped, growing the void that separates the solid electrolyte from residual lithium. Lithium must therefore be extracted from the wall or surface of the MIEC. This excludes dislocation power-law creep as a major kinetic mechanism, since dislocation slip cannot occur in the void, and the residual Li shows little mechanical translation (convection) in our experiments, although slight local sliding cannot be excluded. Therefore, dislocation creep is not the dominant creep mechanism.

To determine the dominant mechanism of Li plating/stripping in MIEC tubules—either interfacial-diffusional Coble creep or bulk diffusional Nabarro–Herring creep—we carried out theoretical calculations (see Methods section ‘Quantitative analysis’). We considered three possible paths for Li diffusion: (a) via an MIEC wall of width w (~10 nm); (b) via the interface between an MIEC wall and Li_{bcc}, with an atomic width of $\delta_{\text{interface}}$ (~2 Å); and (c) via bulk Li_{bcc} of width W (~100 nm). We also considered three canonical MIECs—Li₆C₆, Li₂₂Si₅ and Li₉Al₄. Generally, for the cases when an MIEC is thermodynamically stable against Li_{bcc}, the calculations show that Li diffusion via the interfacial path (b) dominates. This means that the MIEC tubule concept is feasible for Li₉Al₄ and Li₂₂Si₅—or any other electrochemically stable MIEC (for example, CuLi_x, Ni, W) that forms an incoherent interface with Li_{bcc}. In all such cases, the diffusion flux along the 2-Å incoherent interface between the MIEC and the metal, or over the MIEC surface, dominates over flux through the 10-nm MIEC wall itself. In other words, ion transport along the MIEC is dominated by the 2-Å ‘interfacial MIEC channel’, as illustrated in Fig. 1. This greatly widens the range of material choices available for the MIEC, as we can now separate its mechanical function from its electron/ion-transport functions.

Because carbon needs to be lithiated to LiC₆ to become a true MIEC, we introduced ZnO_x during synthesis of the carbon tubules to improve their lithiophilicity, which greatly helps the achievement of uniform-quality MIEC tubules on the first lithiation. We now consider the mechanism of ZnO_x-induced lithiophilicity^{23,24}. On first lithiation, ZnO_x in the MIEC undergoes a conversion/alloying reaction to produce Li₂O, as follows²⁵: $\text{ZnO}_x + (2x + y)\text{Li} = \text{ZnLi}_y + x\text{Li}_2\text{O}$. But it is experimentally difficult to obtain TEM images of the post-formation Li₂O directly: the material is only a few nanometres thick, and located on the inner surfaces of the carbon tubules. We used an alternative method to observe the in situ formed Li₂O, namely imaging the outer surface of the carbon tubules, taking advantage of the homogeneous distribution of ZnO_x across the carbon tubule wall (Supplementary Fig. 2). During Li plating, a crystalline Li₂O layer with a thickness of a few nanometres is observed to be formed along the outer surface of the carbon tubule (Fig. 3a and Supplementary Fig. 18) like a lubricant. Li₂O seems to be mechanically soft, despite its crystallinity, and can also deform by diffusional creep, even at room temperature^{26,27}. If we continue deposition after the interior of the carbon tubule is fully filled with Li, at some point the Li will appear outside the carbon tubule. As shown in Fig. 3b–f and Supplementary Video 9 using dark-field imaging, we observed that after plating through the nanopores, Li first produces a complete wetting, rapidly spreading along the outer surface with zero contact angle up to a distance of 140 nm, before finally pushing downward²⁸. This suggests that the ZnO_x/Li₂O layer on the MIEC surface helps to induce a strong lithiophilicity.

Finally, to demonstrate a centimetre-scale all-solid-state full-cell battery, we constructed an MIEC tubular matrix using about 10¹⁰ cylinders, each with an aspect ratio of several hundred, capped by solid electrolyte (Fig. 4). The counter-electrode is LiFePO₄. To fabricate the tubular MIEC matrix, we first used chemical vapour deposition (CVD) to grow a layer of carbon on the inner surface of free-standing anodic aluminium oxide (AAO) that acted as a template. Next, a layer of Pt was deposited on the bottom of the AAO by sputtering, to act as the current

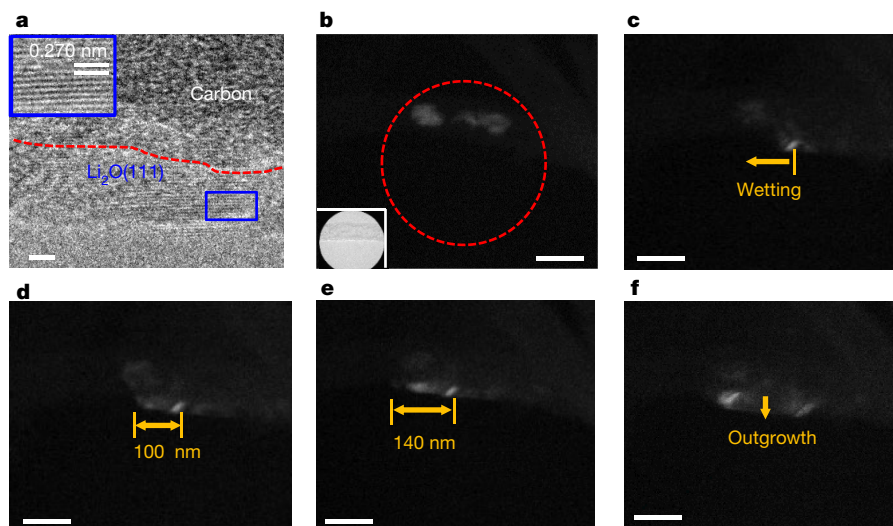


Fig. 3 | Lithiophilicity from ZnO. **a**, HRTEM image of a layer of Li_2O on the outer surface of a carbon tubule. The inset expands the blue-box region, where lattice fringes of Li_2O (111) are seen. **b–f**, Snapshots of dark-field imaging of Li_{bcc} wetting the tubule outer surface as a function of time, with **b** showing the Li_{bcc} already plated inside the carbon tubule, **c** to **e** showing facile wetting on the outside of the tube with spreading distance labelled by yellow arrows, and **f** showing

final pushing downward ('Outgrowth'). For the dark-field imaging, the $(\bar{1}\bar{1}0)$ diffraction beam of the Li crystal shown in **b** is allowed to pass through the objective aperture, and the red dashed circle denotes the selected-area aperture also shown in the inset of **b**. See Supplementary Video 9. Scale bars: **a**, 2 nm; **b–f**, 100 nm.

collector and mechanical support. Then, the AAO was etched away to yield the carbonaceous MIEC tubular matrix²⁹, as shown in Fig. 4a–d. To enhance the lithiophilicity of the carbonaceous MIEC tubular matrix, a 1-nm-thick ZnO layer was deposited onto the surface of the carbon cylinders by atomic layer deposition (Supplementary Fig. 19). This construction, several centimetres in extent and 50 μm thick, sits on the Pt current collector (Supplementary Fig. 20). Indentation tests³⁰ show a hardness of about 65 MPa (Supplementary Fig. 21), which is higher than the internal pressurization limit (see Methods section 'Quantitative analysis'). We then cap the MIEC tubular matrix by a film of PEO-based/LiTFSI solid electrolyte, 50 μm thick. A layer of LiPON about 200 nm thick was pre-deposited into the carbon tubules by sputtering to obstruct the open pores (Supplementary Fig. 22). LiPON has a much

poorer ionic conductivity than PEO-based/LiTFSI solid electrolyte, and approximates as the electron- and Li-ion insulator (ELI) roots shown in yellow in Fig. 1 that affix MIECs to the solid electrolyte. It also prevents inflow of the polymeric solid electrolyte into the MIEC tubules during testing at 55 $^{\circ}\text{C}$. The cathode was constructed from the active material LiFePO_4 (60 wt%), polyethylene oxide (PEO, 20 wt%), LiTFSI (10 wt%) and carbon black (10 wt%). The mass loading is 4–6 mg LiFePO_4 per cm^2 in full cells. In half cells, we use a superabundant Li metal chip (more than 100 \times excess) as the opposite electrode. No (ionic) liquid or gel electrolyte of any kind was used in our centimetre-scale battery experiments. For making the full cell with a small lithium inventory compared to the cathode capacity, we first pre-deposited 1 \times excess Li into the MIEC tubules electrochemically from the half cell.

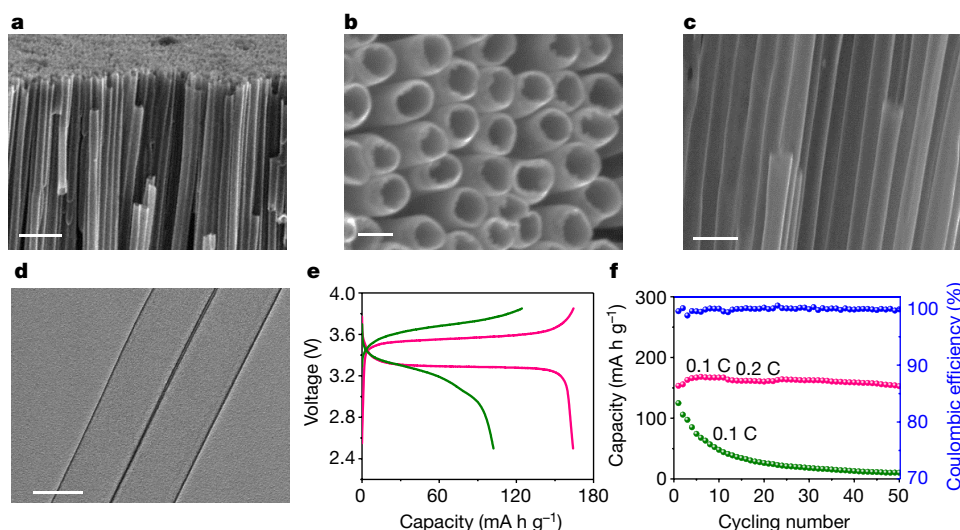


Fig. 4 | Electrochemical performance of scaled-up Li metal cell with about 10^{10} MIEC cylinders. **a–d**, Field emission SEM (FESEM; **a–c**) and TEM (**d**) images of the carbonaceous MIEC tubules. **e, f**, Charge/discharge profiles at 0.1C (**e**) and cycling life (**f**) of the all-solid-state (1 \times excess) Li_{bcc} -pre-deposited MIEC/SE/ LiFePO_4 batteries. The magenta (capacity) and blue (coulombic efficiency)

colours indicate the use of 3D MIEC tubules on Pt foil as a Li host, the discharge capacity of which reaches 164 mA h g^{-1} at 0.1C and 157 mA h g^{-1} at 0.2C, while the green colour indicates the use of 2D carbon-coated Cu foil as a Li host. Scale bars: **a**, 1 μm ; **c**, 500 nm; and **b, d**, 200 nm.

In half-cell tests, we could cycle a large amount of Li with a substantial areal capacity up to 1.5 mA h cm^{-2} . Compared with the control experiment using carbon-coated Cu foil as the Li host, the half cell with the 3D MIEC tubular matrix shows a lower overpotential (39 mV versus 250 mV at 0.125 mA cm^{-2}) and a much higher coulombic efficiency (97.12% versus 74.34% at 0.125 mA cm^{-2}), as well as much better cycling stability (Supplementary Figs. 23, 24). More importantly, in full-cell tests, with only $1\times$ excess Li pre-deposited inside the MIEC tubules, the all-solid-state full cell shows a lower overpotential (0.25 V versus 0.45 V), a higher discharge capacity (164 mA h g^{-1} versus 123 mA h g^{-1}) and a much higher coulombic efficiency (99.83% versus 82.22%) at 0.1 C (Fig. 4e). This full cell shows almost no degradation for more than 50 cycles (Fig. 4f), and the gravimetric capacity of our Li/MIEC composite anode reaches a remarkable value of about 900 mA h g^{-1} . This validates the MIEC architecture for an all-solid-state alkali metal battery, which has been taken from mechanistic concepts to quantitative theory and design to the realm of practice.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1972-y>.

- Porz, L. et al. Mechanism of lithium metal penetration through inorganic solid electrolytes. *Adv. Energy Mater.* **7**, 1701003 (2017).
- Armstrong, R. D., Dickinson, T. & Turner, J. The breakdown of β -alumina ceramic electrolyte. *Electrochim. Acta* **19**, 187–192 (1974).
- Yang, C. et al. Continuous plating/stripping behavior of solid-state lithium metal anode in a 3D ion-conductive framework. *Proc. Natl Acad. Sci. USA* **115**, 3770–3775 (2018).
- Liu, Y. et al. Transforming from planar to three-dimensional lithium with flowable interphase for solid lithium metal batteries. *Sci. Adv.* **3**, eaao0713 (2017).
- Monroe, C. & Newman, J. The impact of elastic deformation on deposition kinetics at lithium/polymer interfaces. *J. Electrochem. Soc.* **152**, A396–A404 (2005).
- Suo, L. et al. Fluorine-donating electrolytes enable highly reversible 5-V-class Li metal batteries. *Proc. Natl Acad. Sci. USA* **115**, 1156–1161 (2018).
- Li, Y. et al. Atomic structure of sensitive battery materials and interfaces revealed by cryo-electron microscopy. *Science* **358**, 506–510 (2017).
- Lu, J., Chen, Z., Pan, F., Cui, Y. & Amine, K. High-performance anode materials for rechargeable lithium-ion batteries. *Electrochem. Energy Rev.* **1**, 35–53 (2018).
- Devaux, D. et al. Failure mode of lithium metal batteries with a block copolymer electrolyte analyzed by X-ray microtomography. *J. Electrochem. Soc.* **162**, A1301–A1309 (2015).
- Harry, K. J., Hallinan, D. T., Parkinson, D. Y., MacDowell, A. A. & Balsara, N. P. Detection of subsurface structures underneath dendrites formed on cycled lithium metal electrodes. *Nat. Mater.* **13**, 69 (2014).
- Maslyn, J. A. et al. Growth of lithium dendrites and globules through a solid block copolymer electrolyte as a function of current density. *J. Phys. Chem. C* **122**, 26797–26804 (2018).
- Harry, K. J., Liao, X., Parkinson, D. Y., Minor, A. M. & Balsara, N. P. Electrochemical deposition and stripping behavior of lithium metal across a rigid block copolymer electrolyte membrane. *J. Electrochem. Soc.* **162**, A2699–A2706 (2015).
- Richards, W. D., Miara, L. J., Wang, Y., Kim, J. C. & Ceder, G. Interface stability in solid-state batteries. *Chem. Mater.* **28**, 266–273 (2016).
- Kim, S. et al. Electrochemically driven mechanical energy harvesting. *Nat. Commun.* **7**, 10146 (2016).
- Jin, C. et al. 3D lithium metal embedded within lithiophilic porous matrix for stable lithium metal batteries. *Nano Energy* **37**, 177–186 (2017).
- Zhao, J. et al. Air-stable and freestanding lithium alloy/graphene foil as an alternative to lithium metal anodes. *Nat. Nanotechnol.* **12**, 993–999 (2017).
- Yan, K. et al. Selective deposition and stable encapsulation of lithium through heterogeneous seed growth. *Nat. Energy* **1**, 16010 (2016).
- Frost, H. & Ashby, M. *Deformation-Mechanism Maps* (Pergamon, 1982).
- Zhu, T. & Li, J. Ultra-strength materials. *Prog. Mater. Sci.* **55**, 710–757 (2010).
- Nitta, N. & Yushin, G. High-capacity anode materials for lithium-ion batteries: choice of elements and structures for active particles. *Part. Part. Syst. Charact.* **31**, 317–336 (2014).
- Chen, Y. et al. Nitrogen-doped carbon for sodium-ion battery anode by self-etching and graphitization of bimetallic MOF-based composite. *Chem* **3**, 152–163 (2017).
- Zheng, H., Liu, Y., Mao, S. X., Wang, J. & Huang, J. Y. Beam-assisted large elongation of in situ formed Li_2O nanowires. *Sci. Rep.* **2**, 542 (2012).
- Liu, Y. et al. Lithium-coated polymeric matrix as a minimum volume-change and dendrite-free lithium metal anode. *Nat. Commun.* **7**, 10992 (2016).
- Li, S. et al. Developing high-performance lithium metal anode in liquid electrolytes: challenges and progress. *Adv. Mater.* **30**, 1706375 (2018).
- Zhang, Y. et al. High-capacity, low-tortuosity, and channel-guided lithium metal anode. *Proc. Natl Acad. Sci. USA* **114**, 3584–3589 (2017).
- Sun, J. et al. Liquid-like pseudoelasticity of sub-10-nm crystalline silver particles. *Nat. Mater.* **13**, 1007–1012 (2014).
- Yang, Y., Kushima, A., Han, W., Xin, H. & Li, J. Liquid-like, self-healing aluminum oxide during deformation at room temperature. *Nano Lett.* **18**, 2492–2497 (2018).
- Kushima, A. et al. Liquid cell transmission electron microscopy observation of lithium metal growth and dissolution: root growth, dead lithium and lithium flotsams. *Nano Energy* **32**, 271–279 (2017).
- Moon, S. et al. Encapsulated monoclinic sulfur for stable cycling of Li-S rechargeable batteries. *Adv. Mater.* **25**, 6547–6553 (2013).
- Cao, G. & Gao, H. Mechanical properties characterization of two-dimensional materials via nanoindentation experiments. *Prog. Mater. Sci.* **103**, 558–595 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Synthesis of carbon hollow tubules

The synthesis of the carbon tubules was similar to that used in our previous work²¹. 1 g of polyacrylonitrile (PAN, Aldrich) and 1.89 g of $\text{Zn}(\text{Ac})_2 \cdot 2\text{H}_2\text{O}$ were dissolved in 30 ml of dimethylformamide (DMF, Aldrich) solvent to obtain the electrospinning solution. A working voltage of 17 kV, a flow rate of 0.05 mm min^{-1} , and an electrospinning distance of 20 cm were used to synthesize the PAN/ $\text{Zn}(\text{Ac})_2$ composite fibres. A layer of zeolitic imidazolate framework (ZIF-8) can be formed on the surface of the composite fibres by adding them to an ethanol solution of 2-methylimidazole (0.65 g, Aldrich). The introduction of a trace amount of cobalt acetate into the composite fibres can promote the graphitization of the carbon tubules. The synthesized core-shell composite fibres were heated at 600–700 °C for 12 h to obtain the hollow carbon tubules with some ZnO_x .

In situ transmission electron microscopy

This was conducted using a JEOL 2010F TEM at 200 kV with a Nano-factory STM/TEM holder³¹. The solid-state nanobattery contains Li metal, solid electrolyte and the prepared carbon tubules with ZnO_x . The Li metal was applied to a tungsten probe in a glove box filled with Ar gas, and the prepared carbon tubules were adhered to half a TEM copper grid by silver conductive epoxy. For a typical example of a soft solid electrolyte, sufficient poly(ethylene oxide) (PEO) and lithium bis(trifluoromethanesulfonyl)imide (LiTFSI) were dissolved in 1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide (ionic liquid). The Li metal on the tungsten probe was capped by the obtained solid electrolyte with a thickness of ~50 µm inside the glove box filled with Ar gas. After loading the battery components into the TEM, the end with solid electrolyte covering the Li metal on the tungsten probe was manipulated to get a contact with the carbon tubules on the TEM copper grid to complete the assembly of a nanobattery. Lithium plating and stripping in the carbon tubules were realized by applying -2 V and +2 V with respect to the lithium metal.

Electron radiation damage control

In the in situ TEM experiments, we reduced electron beam damage as much as possible. Li metal is sensitive to electron beam irradiation in the TEM, owing to elastic and inelastic scattering³². The elastic (electron–nucleus) scattering can lead to sputtering damage, and the inelastic (electron–electron) scattering can cause damage by specimen heating and radiolysis^{7,33}.

In our low-magnification TEM images and videos showing Li plating and stripping inside the tubule, a low electron beam current of around 1.5 mA cm^{-2} was used to minimize beam damage. The images were taken at a slightly underfocused condition to enhance the contrast. We blanked the beam before recording the video, and limited experiment recording time to less than 2 min. The beam was blanked for most of the time while plating and stripping Li, except for some necessary observations. For taking the SAED patterns of Li plated inside the tubule, a broad electron beam with a low electron beam current of 1 mA cm^{-2} was used. We took the patterns as quickly as possible to lower the amount of irradiation damage. For taking the HRTEM image of Li plated inside the tubule, an electron beam current of around 0.3 A cm^{-2} was used. The HRTEM image captured the fresh Li crystal when it first appeared inside the camera field, showing the lattice fringes of (110)_{bcc} planes. The Li lattice fringes remained for several seconds before vanishing owing to electron beam irradiation damage.

The carbon tubules help to reduce irradiation damage when imaging Li_{bcc} inside the tubules (Supplementary Figs. 25–27). As the Li_{bcc} is inside the wall of the tubule, this helps to reduce the sputtering loss. In the case of inelastic scattering, the tubule may also act as a thermal/electron conductor covering the lithium metal, helping to release some heat by electron irradiation. Furthermore, the electrochemical plating

can continually replenish fresh Li_{bcc} in the region under irradiation, which may also help the HRTEM imaging.

We have also carried out in situ TEM experiments with the electron beam blanked. We first set up the nanobattery inside the TEM, placing the selected-area aperture on the still-hollow carbon tubule, and turned on the SAED mode in advance. During these steps, we did not apply any bias potential (the bias potential is required to deposit Li metal inside the carbon tubule). After this preparation, we turned off the electron beam ('blind' condition). With no electron beam present, we applied bias potential for some time to deposit Li metal inside the tubule. We then turned on the electron beam (at a low current density of 1 mA cm^{-2}), and immediately the sharp SAED pattern appeared on the TEM CCD window. The single crystal feature was later identified as the Li_{bcc} phase from its measured lattice constant (Supplementary Fig. 28).

Electron energy-loss spectroscopy (EELS)

The EELS spectra were taken in the STEM mode with a spot size of 1 nm, with a semi-convergence angle of about 5 mrad and a semi-collection angle of about 10 mrad. For the thickness calculation in Supplementary Fig. 8, the absolute log-ratio method was used³⁴, where $\frac{t}{\lambda} = \ln\left(\frac{I_t}{I_0}\right)$ (t stands for thickness, λ stands for effective mean free path, I_t is the intensity integration under the whole EELS spectrum, and I_0 is the intensity integration under the zero loss peak). In addition to the accelerating voltage, semi-convergence and semi-collection angles, to calculate λ the effective atomic number Z_{eff} was also needed. We estimate $Z_{\text{eff}} = 6$ for the carbon tubule before Li plating. After Li plating, both Li and the wall of the carbon tubule existed at the location where we recorded the EELS signal. In this case, we can estimate the rough atomic ratio between Li and C to be 0.56:1 when considering the observed geometry of the tubule, with an inner diameter of ~100 nm and a wall thickness of ~28 nm. Using the formula

$$Z_{\text{eff}} = \frac{\sum_i f_i Z_i^{1.3}}{\sum_i f_i Z_i^{0.3}}$$

we obtain $Z_{\text{eff}} = 5.1$ (after Li plating).

The thicknesses before and after Li plating were thus calculated to be ~68 nm and ~160 nm respectively from the EELS spectra recorded in Supplementary Fig. 8, and the thickness difference (corresponding to the thickness of Li plated) was estimated to be ~92 nm. The background contribution under the edge can be estimated from the pre-edge area, and the K-edge of Li was obtained by background subtraction.

Other characterizations

The synthesized materials were characterized by TEM, high-resolution TEM (HRTEM), field emission scanning electron microscopy (FESEM, FEI Helios 600 Dual Beam FIB), energy-dispersive X-ray spectroscopy (EDX, Oxford) and X-ray photoelectron spectroscopy (XPS, PHI5600).

Synthesis of MIEC 3D electrode

First, the chemical vapour deposition (CVD) method was applied, using 90 sccm C_2H_2 at 640 °C, which grows a layer of carbon onto the inner surface of anodic aluminium oxide (AAO) that acted as the template. Next, a layer of Pt was deposited by sputtering on the bottom of the AAO; it acted as the current conductor and as mechanical support. Then, the AAO was etched to yield the carbonaceous MIEC tubular matrix by employing a 3 M NaOH aqueous solution with a small amount of ethanol added. To enhance the lithiophilicity of the MIEC tubules, a 1-nm-thick ZnO layer was deposited onto the inner surfaces of the MIEC tubular matrix by ALD (atomic layer deposition).

Li/solid electrolyte/MIEC half cell

To avoid the inflow of polymeric solid electrolyte into the MIEC tubules during testing at 55 °C, a layer of LiPON ~200 nm thick was deposited onto the MIEC tubules by sputtering to obstruct the open pores. A

PEO-based/LiTFSI film was used as a typical solid electrolyte (from KISCO Ltd). The 2032 coin cells were prepared by pressing the MIEC tubular matrix onto one side of the solid electrolyte film and a Li metal chip onto the other side. The obtained Li/solid electrolyte/MIEC tubular matrix half cell was tested at current densities of 0.125, 0.25 and 0.5 mA cm⁻². The half cells were cycled a few times to stabilize the interface between solid electrolyte and electrode. The Coulombic efficiency was obtained from the ratio of discharge and charge capacity. For comparison, 2D carbon-coated Cu foil was used to prepare a Li/solid electrolyte/carbon-coated Cu foil cell.

All-solid-state full cell

The LiFePO₄ cathode was constructed from LiFePO₄ powder (the active material, 60 wt%), polyethylene oxide (PEO, 20 wt%), LiTFSI (10 wt%), and carbon black (10 wt%). The mass loading is 4–6 mg LiFePO₄ per cm². We predeposited 1× excess Li into the MIEC tubules from the half cell. The 2032 coin cells were prepared with Li-deposited MIEC tubules as the anode, the LiFePO₄ electrode as the cathode and solid electrolyte in an Ar-filled glove box. The all-solid-state battery was tested at 55 °C with a LAND battery tester between 2.5 V and 3.85 V. We also predeposited 1× excess Li onto the carbon-coated Cu foil of the control battery before the full cell testing.

Quantitative analysis

Internal gas pressure accommodation. For mechanical stability, it is in practice difficult to construct a vacuum-filled tubular matrix, so we will assume that initially we have an inert gas phase in the white region in Fig. 1 with $P_{\text{gas}} = 1$ atm. The gas-tightness of the solid electrolyte layer must be guaranteed, because otherwise Li metal will easily plate or flow through the solid electrolyte, shorting to the cathode. Thus, when Li metal is deposited inside a tubule, the gas phase must be compressed. If the current collector (say Cu) and the MIEC walls are also hermetically sealed, then local P_{gas} will increase as more and more Li metal is deposited inside, up to possibly tens of atmospheres (a few MPa) if the compression ratio is something like 10×. The creeping Li metal can act as a piston, as we have seen from the Nernst equation that P_{LiMetal} can easily reach hundreds of MPa. However, owing to unavoidable heterogeneities, the amount of Li metal deposited may not be the same between adjacent cylinders, and this will cause a pressure difference, ΔP_{gas} , between adjacent cylinders that can bend the MIEC wall. If the MIEC wall (red region in Fig. 1) is not mechanically ductile enough, then at a certain point a cell may burst. For this reason, it is better for the MIEC wall to be permeable, so P_{gas} can then equilibrate from cell to cell. Then the internal pressure will be more homogeneously distributed, ensuring that the left chamber in Fig. 1 will not expand and crush the right chamber by bending the wall.

Geometric design. While the in situ TEM experiments give us confidence that MIEC electrochemical cells work at the ‘single cylinder’ or ‘few-cylinders’ level, transport and mechanical durability issues will determine how well the cell will work in practice at cm × cm scale, with a massive number ($\sim 10^{10}$) of parallel cylinders. The typical areal capacity Q and current density $J \equiv dQ/dt$ demanded by industrial applications are of the order 3 mA h cm⁻² and 3 mA cm⁻², respectively. Typical overpotentials U of lithium-metal-containing anodes (versus Li⁺/Li) are of the order of 50 mV. With unavoidable heterogeneities among the $\sim 10^{10}$ cylinders, transport/reaction limitations may vary from location to location. With P_{LiMetal} in MPa and U in V, we have $\max P_{\text{LiMetal}} = 7,410U$, so for $U = 50$ mV, $\max P_{\text{LiMetal}} = 370$ MPa: the higher the overpotential, the larger $\max P_{\text{LiMetal}}$, and the more severe the local mechanical degradations can be. We cannot allow the overpotential U , a global quantity, to rise too high; but U is still responsible for driving a global average current density J . This means the average transport conductance should be better than $\sim 3 \text{ mA cm}^{-2} / 50 \text{ mV} = 0.06 \text{ S cm}^{-2}$ as an order-of-magnitude estimate, otherwise the requisite pressure might be too high

and the MIEC tubules may burst somewhere. The effective transport conductance of the tubular matrix is $(\kappa_{\text{MIEC}}/h) \times w/(w+W)$, where κ_{MIEC} (in S cm⁻¹) is an effective Li conductivity, and $w/(w+W)$ is the fill factor by MIEC (assuming straight pores and tortuosity = 1). In order to get $Q \approx 3 \text{ mA h cm}^{-2}$, h needs to be at least $\sim 20 \mu\text{m}$, taking into account the inert host volume (see Supplementary Fig. 29 for calculated capacity with the tubular matrix geometry). So we get an effective longitudinal transport requirement:

$$\kappa_{\text{MIEC}} \times w/(w+W) > 0.06 \text{ S cm}^{-2} \times 20 \mu\text{m} = 0.12 \text{ mS cm}^{-1} \quad (1)$$

For MIEC, we have bulk contribution

$$\kappa_{\text{MIEC}}^{\text{bulk}} \approx e^2 c_{\text{Li}} D_{\text{Li}}^{\text{bulk}} / k_B T \quad (2)$$

where c_{Li} (in cm⁻³) is the Li atom concentration, and $D_{\text{Li}}^{\text{bulk}}$ is the tracer diffusivity of Li atoms in bulk MIEC. We should recognize, however, that interfacial diffusion might be significant or even dominant with 100-nm-sized MIEC cylinders, as there can be fast diffusion paths of width $\delta_{\text{interface}}$ (typically taken to be 2 Å) at the MIEC/Li_{bcc} incoherent phase boundary (red/grey interface in Fig. 1) or surface (red/white interface in Fig. 1), in which case we need to correct κ_{MIEC} by the following size-dependent factor:

$$\kappa_{\text{MIEC}} = \kappa_{\text{MIEC}}^{\text{bulk}} \times (1 + 2D_{\text{Li}}^{\text{interface}} \delta_{\text{interface}} / D_{\text{Li}}^{\text{bulk}} w) \quad (3)$$

With bulk diffusivity data culled from table 2 of ref.²⁰, we see that among the three canonical MIECs—LiC₆ ($c_{\text{Li}} = 1.65 \times 10^{22} \text{ cm}^{-3}$, optimistic $D_{\text{Li}}^{\text{bulk}} \approx 10^{-7} \text{ cm}^2 \text{ s}^{-1}$), Li₂₂Si₅ ($c_{\text{Li}} = 5.3 \times 10^{22} \text{ cm}^{-3}$, optimistic $D_{\text{Li}}^{\text{bulk}} \approx 10^{-11} \text{ cm}^2 \text{ s}^{-1}$) and Li₉Al₄ ($c_{\text{Li}} = 4 \times 10^{22} \text{ cm}^{-3}$, optimistic $D_{\text{Li}}^{\text{bulk}} \approx 10^{-9} \text{ cm}^2 \text{ s}^{-1}$)—it looks likely that LiC₆ has the largest $c_{\text{Li}} D_{\text{Li}}^{\text{bulk}}$. Putting the values into equation (2), $\kappa_{\text{MIEC}}^{\text{bulk}}(\text{LiC}_6) \approx 0.01 \text{ S cm}^{-1}$. However, there is large uncertainty in the diffusivity data, so a more conservative estimate might be $D_{\text{Li}}^{\text{bulk}}(\text{LiC}_6) \approx 10^{-8} \text{ cm}^2 \text{ s}^{-1}$, $\kappa_{\text{MIEC}}^{\text{bulk}}(\text{LiC}_6) \approx 1 \text{ mS cm}^{-1}$. Thus, the minimum MIEC fill factor for LiC₆ is

$$w_{\text{min}}/(w_{\text{min}} + W) = \frac{0.12 \text{ mS cm}^{-1}}{\kappa_{\text{MIEC}}} \approx 0.1$$

and so if $W \approx 100 \text{ nm}$, one should have minimally $w = w_{\text{min}} \approx 10 \text{ nm}$. This wall thickness happens to also make sense from a mechanical robustness requirement viewpoint. Coincidentally, this geometry is quite close to that of our carbon tubule experiment. The design above is consistent with the fact that graphite or hard carbon anodes used in lithium-ion batteries (LIB) have a film thickness of the order of 100 μm, and the film is known to be able to support a current density of $\sim 3 \text{ mA cm}^{-2}$ with an overpotential of $\sim 50 \text{ mV}$. Indeed, referencing to an industrial LIB graphite anode is apt here, because we know they work near the borderline as an anode in charging: if the current density is significantly higher than $\sim 3 \text{ mA cm}^{-2}$, then the local potential would drop below 0 V versus Li⁺/Li, and Li_{bcc} would precipitate out, which is a substantial problem for LIB cycle life and safety with liquid electrolyte. Here, we are proposing to turn the problem on its head. We want the Li metal to ‘spill out’ of the MIEC, but in a controlled fashion, inside the internal tubular cells within a reserved space capped by ELI and solid electrolyte, without excessive P_{LiMetal} build-up and cracking of the solid electrolyte, and without any fresh SEI production (since the expanding/shrinking parts are in contact with MIEC and will stop at ELI, which are both electrochemically absolutely stable against Li metal, so no side reactions are possible electrochemically). Then we only need to ensure mechanical integrity of this 3D solid structure of open-pore MIECs rooted in solid electrolyte via ELI.

If there were no interfacial diffusion contribution, Li₉Al₄ might be a borderline case, with $\kappa_{\text{MIEC}}(\text{Li}_9\text{Al}_4) \approx 0.25 \text{ mS cm}^{-1}$ from equation (2), thus requiring an excessively large MIEC fill factor of

$$w_{\min}/(w_{\min} + W) = \frac{0.12 \text{ mS cm}^{-1}}{\kappa_{\text{MIEC}}} \approx 0.5$$

and requiring $w_{\min} \approx W \approx 100 \text{ nm}$. Such a large fill factor is unlikely to be competitive against a graphite anode. Lastly, the bulk Li diffusivity value ($D_{\text{Li}}^{\text{bulk}} \approx 10^{-11} \text{ cm}^2 \text{ s}^{-1}$) for $\text{Li}_{22}\text{Si}_5$ is totally unworkable, because $\kappa_{\text{MIEC}}(\text{Li}_{22}\text{Si}_5) \approx 0.003 \text{ mS cm}^{-1}$, and we cannot satisfy the transport requirement, equation (1). We conclude therefore that if bulk diffusion alone operates in the MIEC, $D_{\text{Li}}^{\text{bulk}} \approx 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ would be workable, $D_{\text{Li}}^{\text{bulk}} \approx 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ would be difficult, and anything lower would be impossible.

Experimentally, when stripping Li metal (Fig. 2j–l, Supplementary Fig. 17 and Supplementary Video 4), we can sometimes create a void plug that grows between the residual Li metal and the solid electrolyte. Yet this gap does not prevent the Li metal from being further stripped in the experiment, growing the void that separates solid electrolyte from the residual lithium. Li metal must therefore flow out from the MIEC wall/surface. This then excludes dislocation power-law creep as a major kinetic mechanism, since dislocation slip cannot occur in the void, and the residual Li metal shows very little mechanical translation (convection) in our experiments, although slight local sliding cannot be excluded. Based on our in situ TEM observations, therefore, the Li metal must be in the Coble creep regime. However, this does not determine whether the Li is transported along the MIEC interior of width w , or along the MIEC/Li metal interface (Supplementary Fig. 4 case) or over the MIEC surface (Fig. 2j–l case) of width $\delta_{\text{interface}}$, and then plated to the tip of the Li metal via Li metal surface diffusion, as illustrated in Fig. 1. A theoretical bound is necessary. According to NMR measurements³⁵, bulk b.c.c. Li metal has $D_{\text{Li}}^{\text{bulk}} \approx 4 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$ at room temperature, which we know from the calculations above is two orders of magnitude too sluggish to support the observed Li metal kinetics. For surface diffusivity of Li on b.c.c. Li metal, the empirical formula³⁶

$$D_{\text{Li}}^{\text{surface}} = 0.014 \exp(-6.54 T_{\text{M}}/T) \text{ cm}^2 \text{ s}^{-1} \quad (4)$$

has been verified to work quite well for monatomic metals. For instance, Sn, another low-melting-point metal ($T_{\text{M}} = 232 \text{ }^{\circ}\text{C}$), was found to have surface diffusivity $D_{\text{Sn}}^{\text{surface}} \approx 1 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$ at room temperature by direct mechanical creep deformation experiments³⁷, while equation (4) predicts $2 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$. Equation (4) predicts $D_{\text{Li}}^{\text{surface}} = 7 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$ in b.c.c. Li at room temperature. This is $70\times$ larger than that of $D_{\text{Li}}^{\text{bulk}} \approx 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ in LiC_6 . The geometry factor $2\delta_{\text{interface}}/w$, on the other hand, is of the order of $4 \text{ }^{\circ}\text{A}/10 \text{ nm} = 1/25$. Thus, if one takes an optimistic estimate that $D_{\text{Li}}^{\text{interface}} \approx D_{\text{Li}}^{\text{surface}}$, then the interfacial diffusion contribution can be $3\times$ that of the bulk MIEC diffusional contribution even for LiC_6 . The MIEC/Li metal phase boundary has a lower atomic free volume than the free Li metal surface, so we expect $D_{\text{Li}}^{\text{interface}}$ could be a factor of a few smaller than $D_{\text{Li}}^{\text{surface}} = 7 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$. Experimental diffusivity data for metals³⁸ suggest that $D_{\text{Li}}^{\text{interface}} \approx 2 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$. Thus, $D_{\text{Li}}^{\text{interface}}$ will definitely dominate over bulk MIEC diffusion for Li_9Al_4 and $\text{Li}_{22}\text{Si}_5$, as the ratio $D_{\text{Li}}^{\text{interface}}/D_{\text{Li}}^{\text{bulk}}$ (200 for Li_9Al_4 and 20,000 for

$\text{Li}_{22}\text{Si}_5$) easily overwhelms the geometric factor $2\delta_{\text{interface}}/w$ ($1/25$ for $w = 10 \text{ nm}$). The bulk MIEC contribution can thus be ignored for the electrochemical design, and regardless of MIEC choices we predict an effective $\kappa_{\text{MIEC}} \approx 1 \text{ mS cm}^{-1}$, which would satisfy the longitudinal transport requirement, equation (1), for an MIEC fill factor of $w/(w + W) = 0.1$. This predicts that the MIEC tubule concept actually becomes feasible even for Li_9Al_4 and $\text{Li}_{22}\text{Si}_5$ or any other electrochemically stable MIEC, since diffusion flux along the $\delta_{\text{interface}} \approx 2 \text{ }^{\circ}\text{A}$ MIEC/metal incoherent interface or the MIEC surface dominates over the 10 nm MIEC itself. This recognition greatly liberates the MIEC material selection choices, as we can now separate its mechanical function from its ion-transport function. In other words, ion transport along the MIEC is dominated by an ‘interfacial MIEC channel’ along $\delta_{\text{interface}}$, as illustrated in Fig. 1.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

- Yuan, Y., Amine, K., Lu, J. & Shahbazian-Yassar, R. Understanding materials challenges for rechargeable ion batteries with in situ transmission electron microscopy. *Nat. Commun.* **8**, 15806 (2017).
- Egerton, R. F., Li, P. & Malac, M. Radiation damage in the TEM and SEM. *Micron* **35**, 399–409 (2004).
- Wang, X. et al. New insights on the structure of electrochemically deposited lithium metal and its solid electrolyte interphases via cryogenic TEM. *Nano Lett.* **17**, 7606–7612 (2017).
- Malis, T., Cheng, S. C. & Egerton, R. F. EELS log-ratio technique for specimen-thickness measurement in the TEM. *J. Electron Microsc. Tech.* **8**, 193–200 (1988).
- Mali, M., Roos, J., Sonderegger, M., Brinkmann, D. & Heitjans, P. ⁶Li and ⁷Li diffusion coefficients in solid lithium measured by the NMR pulsed field gradient technique. *J. Phys. F* **18**, 403 (1988).
- Xie, D.-G. et al. In situ study of the initiation of hydrogen bubbles at the aluminium metal/oxide interface. *Nat. Mater.* **14**, 899–903 (2015).
- Tian, L., Li, J., Sun, J., Ma, E. & Shan, Z.-W. Visualizing size-dependent deformation mechanism transition in Sn. *Sci. Rep.* **3**, 2113 (2013).
- Gjostein, N. A. *Diffusion* (ASM, 1973).

Acknowledgements We acknowledge support by the Department of Energy, Basic Energy Sciences under award number DE-SC0002633 (‘Chemomechanics of far-from-equilibrium interfaces’), and by NSF ECCS-1610806. We thank KISCO Ltd for providing the PEO-based/LITFSI solid electrolyte film.

Author contributions The experiments were conceived and designed by Y.C., Z.W. and J.L.; Y.C. and Z.W. performed the in situ TEM experiments, TEM imaging analysis, materials characterization, and the electrochemical performance assessments; Y.C., Z.W., X.L. and X.Y. synthesized the materials; Y. L., N.W. and J.B.G. helped with the electrochemical characterization; S.Y.K. and Y.-W.M. performed the nanoindentations; Y.C., Z.W. and J.L. wrote the paper; Y.-W.M., Z.W., X.L., X.Y., C.W., W.X., D.Y., F.Y., A.K., G.Z., H.H., J.B.G. and J. L. analysed the data, discussed the results and commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1972-y>.

Correspondence and requests for materials should be addressed to J.L.

Peer review information Nature thanks Werner Sitte and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Hidden diversity of vacancy networks in Prussian blue analogues

<https://doi.org/10.1038/s41586-020-1980-y>

Received: 22 August 2019

Accepted: 23 December 2019

Published online: 12 February 2020

Arkadiy Simonov^{1,2}, Trees De Baerdemaeker^{1,3}, Hanna L. B. Boström^{1,4},
María Laura Ríos Gómez^{5,6}, Harry J. Gray¹, Dmitry Chernyshov⁷, Alexey Bosak⁸,
Hans-Beat Bürgi^{9,10} & Andrew L. Goodwin^{1*}

Prussian blue analogues (PBAs) are a diverse family of microporous inorganic solids, known for their gas storage ability¹, metal-ion immobilization², proton conduction³, and stimuli-dependent magnetic^{4,5}, electronic⁶ and optical⁷ properties. This family of materials includes the double-metal cyanide catalysts^{8,9} and the hexacyanoferrate/hexacyanomanganate battery materials^{10,11}. Central to the various physical properties of PBAs is their ability to reversibly transport mass, a process enabled by structural vacancies. Conventionally presumed to be random^{12,13}, vacancy arrangements are crucial because they control micropore-network characteristics, and hence the diffusivity and adsorption profiles^{14,15}. The long-standing obstacle to characterizing the vacancy networks of PBAs is the inaccessibility of single crystals¹⁶. Here we report the growth of single crystals of various PBAs and the measurement and interpretation of their X-ray diffuse scattering patterns. We identify a diversity of non-random vacancy arrangements that is hidden from conventional crystallographic powder analysis. Moreover, we explain this unexpected phase complexity in terms of a simple microscopic model that is based on local rules of electroneutrality and centrosymmetry. The hidden phase boundaries that emerge demarcate vacancy-network polymorphs with very different micropore characteristics. Our results establish a foundation for correlated defect engineering in PBAs as a means of controlling storage capacity, anisotropy and transport efficiency.

The true crystal structures of PBAs—and of Prussian blue itself—have long posed a difficult and important problem in solid-state chemistry because their ostensibly simple powder diffraction patterns (Fig. 1a) belie a remarkable complexity at the atomic scale^{17–19}. The common parent structure is based on the cubic lattice and corresponds to the idealized composition $M[M'(CN)_6]_x$. Atoms of type M and M' (usually transition-metal cations) occupy alternate lattice vertices and are octahedrally coordinated by bridging cyanide ions (CN^-) at the lattice edges (Fig. 1b, left). There is a close parallel to the double perovskite structure; indeed the considerations of covalency and octahedral coordination that stabilize perovskites among oxide ceramics also favour this same architecture for transition-metal cyanides, accounting for the chemical diversity of PBAs²⁰. Charge balance requires that the formal oxidation states of M and M' sum to six, as in $Cd^{II}[Pd^{IV}(CN)_6]_2$ (ref. 21).

Prussian blue itself is a mixed-valence cyanide of iron in its 2+ and 3+ oxidation states^{22,23}, and so its composition cannot respect this oxidation-state-sum rule. Instead the rule is circumvented by vacancies: the composition is well approximated by the formula $Fe^{III}[Fe^{II}(CN)_6]_{3/4}\square_{1/4}\cdot xH_2O$, where the symbol \square represents a $[Fe^{II}(CN)_6]^{4-}$ vacancy¹⁸. Vacancies are usually filled by water molecules, which complete the coordination sphere of the neighbouring M cations¹²; we use

the term 'vacancy' to encompass the possible occupancy of the M' site with water. Each vacancy gives rise to a micropore with an effective diameter of approximately 8.5 Å that exceeds the distance between neighbouring M' sites ($a/\sqrt{2} \approx 7.2$ Å)²⁴. Hence a pair of neighbouring vacancies, if present, connects to form a larger micropore¹. A random vacancy distribution would imply bulk microporosity, because the vacancy fraction exceeds the percolation threshold for the face-centred cubic (fcc) M' sublattice (about 0.20)²⁵. But Prussian blue is not microporous: single-crystal X-ray diffraction has shown that vacancies tend to avoid one another by adopting a specific ordered arrangement (Fig. 1b, centre)¹⁸. A vacancy fraction of 1/4 is the greatest that can support complete vacancy isolation.

PBAs with a nominal composition of $M^{II}[M^{III}(CN)_6]_{2/3}\square_{1/3}\cdot xH_2O$ (hereafter $M[M']$) contain an even higher fraction of M'-site vacancies^{12,20,26}. Hence geometry dictates that these vacancies—whatever their distribution—must form connected neighbour pairs (Fig. 1b, right). The existence and nature of any extended micropore network that then develops depends on longer-range vacancy correlations. The collective micropore structure of PBAs is remarkably poorly understood, despite the relevance of mass transport to the many important properties of the family^{1,11}. We do know the following: adsorption isotherm

¹Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford, UK. ²Laboratory for Multifunctional Ferroic Materials, Department of Materials, ETH Zürich, Zürich, Switzerland. ³Centre for Surface Chemistry and Catalysis, KU Leuven, Leuven, Belgium. ⁴Department of Chemistry, Uppsala University, Uppsala, Sweden. ⁵Departamento de Polímeros, Instituto de Investigaciones en Materiales, Universidad Nacional Autónoma de México, Mexico City, Mexico. ⁶Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, UK. ⁷Swiss–Norwegian Beam Lines, European Synchrotron Radiation Facility, Grenoble, France. ⁸European Synchrotron Radiation Facility, Grenoble, France. ⁹Department of Chemistry, University of Zürich, Zürich, Switzerland. ¹⁰Department of Chemistry and Biochemistry, University of Berne, Bern, Switzerland. *e-mail: andrew.goodwin@chem.ox.ac.uk

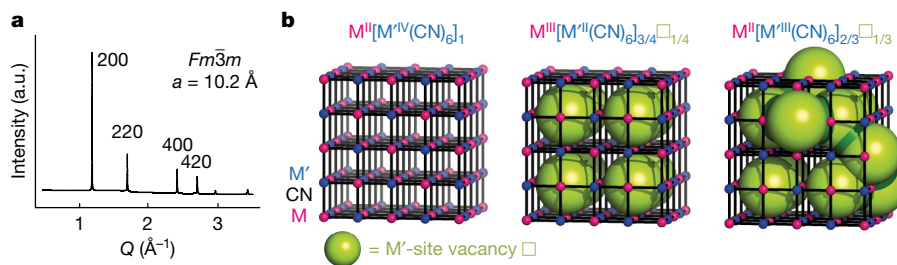


Fig. 1 | Structure of PBAs. **a**, PXRD pattern of Mn[Co], typical for PBAs. **b**, The parent structure type (left) comprises interpenetrating fcc arrays of M and M' cations (pink and blue spheres, respectively), bridged by cyanide ions (black rods). In Prussian blue (centre), one-quarter of the M' sites are vacant, creating isolated micropores (green spheres) that are usually occupied by water. In

PBAs (right), one-third of the M' sites are vacant. There are now sufficiently many vacancies that neighbouring pores must connect (dark green collars) to give an extended micropore network. The characteristics of this network depend on vacancy correlations.

measurements reflect a diversity of pore characteristics across PBAs^{26,27}; solid-state ¹¹³Cd NMR measurements have evidenced non-statistical vacancy distributions in Cd[Fe_xCo_{1-x}] (ref. ²⁸); weak primitive superlattice reflections have been observed only sometimes in powder X-ray diffraction (PXRD) patterns—their presence has usually been interpreted as evidence for (partial) Prussian-blue-type vacancy order¹⁶; high-resolution transmission electron microscopy has revealed vacancy chains in some copper-containing PBAs and their absence in zinc-containing samples²⁹; and in the only existing single-crystal diffraction study of a PBA (namely Mn[Mn]), structured diffuse scattering was observed and interpreted in terms of Warren–Cowley correlation parameters^{30,31}. Taken together, these observations suggest that vacancy distributions are unlikely to be random, and that there must be substantial variability in the pore networks of different PBAs.

In this study, we have characterized vacancy correlations in a range of PBAs by growing single crystals, measuring their X-ray diffuse scattering patterns, and interpreting these patterns via a three-dimensional difference pair distribution function (3D-ΔPDF) analysis and Monte Carlo simulations.

For every crystal we investigated, the corresponding X-ray diffraction pattern contained weak but highly structured diffuse scattering, which is the hallmark of strongly correlated disorder³². Representative (*h**k*0) cuts of our diffuse scattering patterns are shown for a selection of PBAs in Fig. 2, where we also include the single-crystal diffuse scattering pattern of Mn[Mn]—the only other such pattern ever reported for a PBA^{30,31}. The inverse Fourier transform of the normalized diffuse scattering function yields the 3D-ΔPDF³³. The form of all of our 3D-ΔPDFs is

well described by a convolution of the contribution from an individual [M^{III}(CN)₆]³⁻ anion together with an occupational correlation function. Hence the diffuse scattering we observe arises from correlations in [M^{III}(CN)₆]³⁻ occupancy instead of from any alkali cation or solvent inclusion. The scattering is also predominantly elastic, because the 3D-ΔPDF is dominated by occupational correlations and not the signature of cooperative displacements. In PXRD measurements, orientational averaging conceals the diffuse scattering within the background or causes it to resemble primitive superlattice reflections³¹; it is in this sense that the vacancy correlations from which the diffuse scattering arises are ‘hidden’.

We find a surprising diversity of diffuse scattering patterns among the studied PBAs. This is true even for crystals with the same nominal composition but grown separately (the example in Fig. 2 is a pair of Mn[Co] crystals grown in different media). So our experimental data unambiguously show that the vacancies in PBAs are distributed in a highly non-random manner, and that these distributions can be fundamentally different for different samples.

It remains to show how we might understand this diversity, and what the implications are for mass transport in PBAs. To do so, we have developed a very simple vacancy interaction model that is nevertheless capable of rationalizing the various experimental diffuse scattering patterns. Monte Carlo simulations driven by this set of interactions generate representative pore-network configurations for each phase that can then be used to determine physical properties of relevance to mass transport and storage. Our model contains just two components, each based on simple crystal-chemical considerations. The first favours a uniform vacancy distribution, such that for each M site four of the six neighbouring M' sites are occupied and two are vacant. This contribution reflects Pauling's ‘electroneutrality’ principle³⁴. The second component favours locally centrosymmetric arrangements, which we expect to have greater or lesser importance depending on the M-site chemistry.

Formally, we represent the Monte Carlo energy by

$$E = \sum_{\mathbf{r} \in \{\text{M}\}} \left[J_1 \left(4 - \sum_{\mathbf{r}' \in \frac{1}{2}(100)} e_{\mathbf{r}+\mathbf{r}'} \right)^2 + \frac{J_2}{2} \sum_{\mathbf{r}' \in \frac{1}{2}(100)} (e_{\mathbf{r}+\mathbf{r}'} - e_{\mathbf{r}-\mathbf{r}'})^2 \right] \quad (1)$$

where the sum is taken over all M sites at positions \mathbf{r} , with the neighbouring M'-site states $e_{\mathbf{r} \pm \mathbf{r}'} = 0$ (vacant) or 1 (present), and $J_1, J_2 > 0$ quantifying the strength of the electroneutrality and centrosymmetry terms, respectively. The occupancy fraction $\langle e \rangle = 2/3$. The quadratic form of the electroneutrality component comes from the leading term in the series expansion in local charge at the M site. We performed Monte Carlo simulations for a range of J_1/J_2 ratios and effective temperatures $T^* = T/J_2$. Our results are shown in Fig. 3a, represented in terms of the single-crystal X-ray diffuse scattering patterns calculated from an

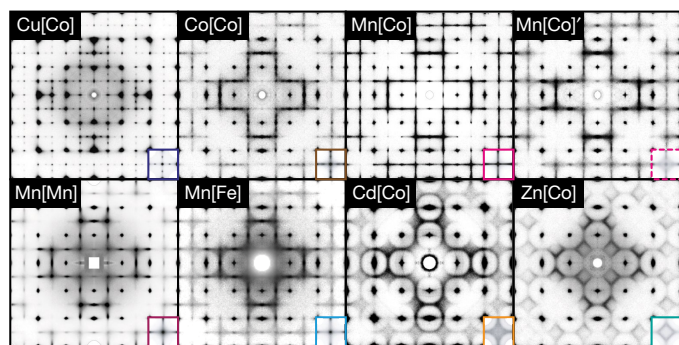


Fig. 2 | Single-crystal diffuse scattering from PBAs. Reconstructed (*h**k*0) scattering planes are shown here for eight PBA samples ($-6 < |h|, |k| < 6$). The data for Mn[Mn] are those reported in ref. ³¹. At the bottom-right corner of each panel is the diffuse scattering pattern averaged over all squares with $\delta h, \delta k = 2$ in the (*h**k*0) scattering plane. Intensities near the Bragg positions with even *h*, *k* in the corners of the squares have been removed. Note the fundamental difference in information content of these single-crystal data relative to PXRD traces of the same materials; compare with Fig. 1a.

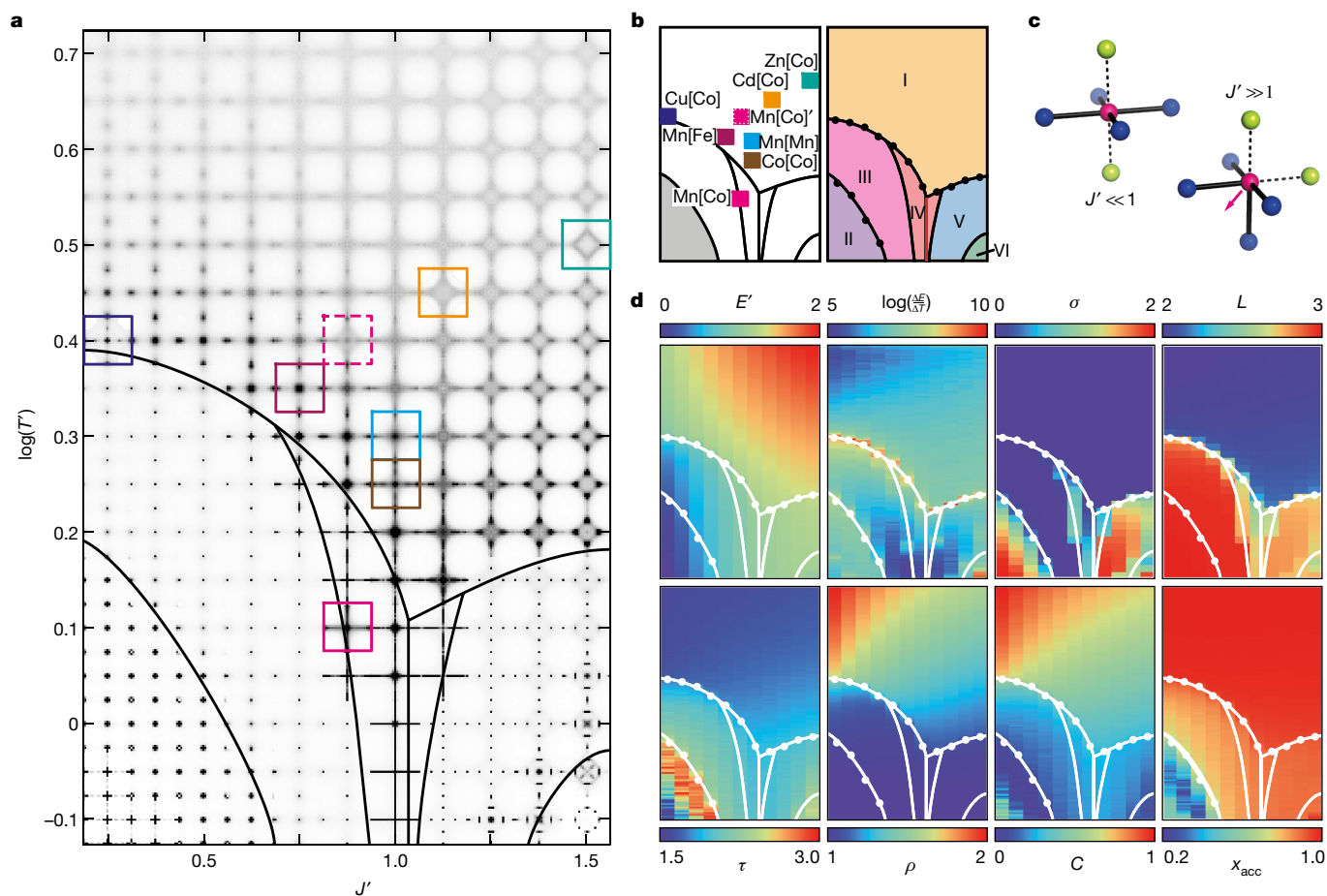


Fig. 3 | Vacancy network phase diagram. **a**, Monte Carlo diffuse scattering map with experimental plane-averaged scattering superimposed (squares). **b**, Distribution of PBAs (left) and vacancy polymorphs (I–VI) demarcated by Monte Carlo specific-heat anomalies (black circles) and a morphotropic phase boundary (red line)⁵¹. Lines are guides to the eye. **c**, Centrosymmetric and pseudotetrahedral M-site geometries. **d**, Thermodynamic and micropore

network characteristics: normalized Monte Carlo energy E' ; Monte Carlo energy gradient $\log(\Delta E/\Delta T)$; anisotropy $\sigma = \sum (I - \bar{I})^2$, where I and \bar{I} are the diffuse scattering intensities before and after Laue symmetrisation; scattering localization $L = \log[\Sigma(I^2)/(\Sigma I)^2]$; surface-accessible vacancy fraction x_{acc} ; conductance C ; vacancy-neighbour pairs per formula unit ρ ; and tortuosity τ .

ensemble of 40 Monte Carlo configurations for each point across an evenly distributed mesh of J' , $\log(T')$.

The phase behaviour given by this simple Monte Carlo model is remarkable for a number of reasons. Clearly the form of the diffuse scattering—and, as we will come to see, of the vacancy-network topology—is an extremely sensitive function of J' and T' . This observation mirrors our experimental results: namely, that small variations in synthesis conditions or PBA composition strongly affect the diffuse scattering. Such sensitivity arises because the two interaction terms of electroneutrality and centrosymmetry operate in tension: they are resolvable when the vacancy fraction is $1/4$ (giving the ordered Prussian blue vacancy arrangement shown in Fig. 1b; compare with sample I in ref.¹⁸), but become frustrated as the vacancy fraction increases. Hence the crystal-chemical considerations embedded in equation (1) drive an unexpectedly complex configurational landscape for PBAs with an M'-site vacancy fraction of $1/3$. We note the parallel to geometric frustration in relaxor ferroelectrics (for example, $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$) and relaxor ferromagnets (for example, $\text{La}(\text{Sb}_{1/3}\text{Ni}_{2/3})\text{O}_3$), where the problem of 1:2 decoration of the fcc lattice is also central^{35,36}.

The experimental diffuse scattering patterns given in Fig. 2 are well approximated by our Monte Carlo simulations at different values of J' and T' (Fig. 3a, b). The implication is that electroneutrality and centrosymmetry are alone sufficient to account for the basic form and diversity of the diffuse scattering patterns that are observed experimentally.

But what determines J' and T' for a given system? PBAs with Jahn–Teller-active M-site cations (such as $\text{Cu}[\text{Co}]$) correspond to smaller values of J' , which is sensible because crystal field effects²⁰ must increase the relative importance of the J_2 term. By contrast, crystal-field-inactive M-site cations correspond to larger J' ; the larger values for $\text{Zn}[\text{Co}]$ and $\text{Cd}[\text{Co}]$ probably reflect the empirical propensity of Zn and Cd to adopt acentric coordination geometries in their pseudobinary cyanides^{37,38} and rhombohedral PBAs³⁹ (Fig. 3c). So PBA composition controls J' , with M-site chemistry more important than that of the M' site. Solid solutions will probably span the range of J' values bounded by the corresponding endmembers, which in the case of Cu/Zn mixtures renders most of J' space accessible synthetically. The effective Monte Carlo temperature T' appears not to be driven by composition but reflects instead the precursor concentration and crystal growth rate (high $T' \equiv$ rapid precipitation and/or high oversaturation). Our $\text{Mn}[\text{Co}]$ and $\text{Mn}[\text{Co}]'$ samples are associated with similar J' but different T' , with the lower T' value for the slower-grown sample (by gel diffusion). By reducing the synthesis temperature and/or the precursor concentrations, it may prove possible to access $\log T'$ values lower than those we report here. So, from a synthetic viewpoint, there is genuine scope for navigating much of the J' and T' space through judicious choice of the PBA chemistry (J') and synthesis approach (T').

Just as the calculated diffuse scattering patterns are unexpectedly diverse for our Monte Carlo configurations, so too are the

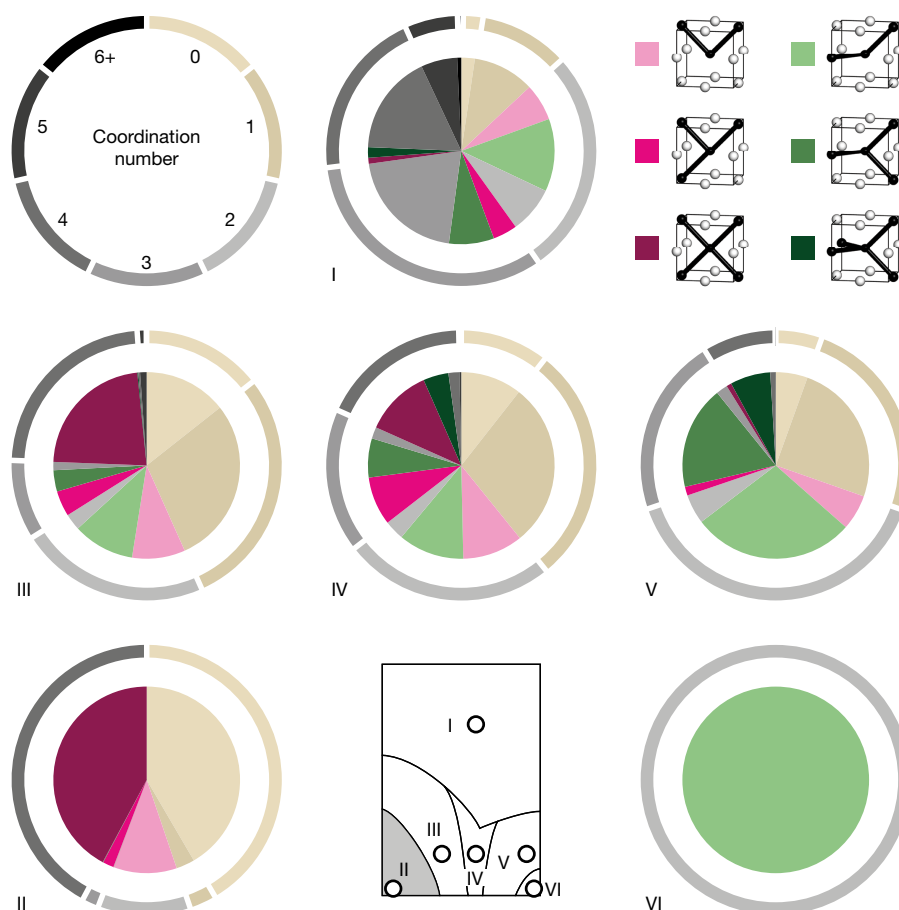


Fig. 4 | Statistical properties of micropore networks. Network coordination number and geometry distributions are given as interior and exterior pie charts, respectively, for representative configurations drawn across the phase diagram shown in Fig. 3. For each coordination number ≥ 2 , coordination geometries related to square-planar networks are shaded pink; those related to

corresponding vacancy-network structures. Despite their considerable disorder, these networks have meaningfully different physical characteristics that we discuss in greater detail below. At the simplest level, different configurations have vacancy networks with different coordination number and geometry distributions (Fig. 4 and Extended Data Fig. 1). Low values of J' give networks dominated by square-planar nodes; at large J' , we find low-dimensional motifs based on 120° zig-zag chains instead. High T' favours a greater diversity of network geometries and low effective temperatures stabilize uniform vacancy networks and/or phase segregation.

Collectively, the different scattering patterns and micropore geometries identify previously unknown phase domains of distinct vacancy-network polymorphs, the boundaries between which are hidden from conventional PXRD analysis (Fig. 3b). These boundaries emerge from our Monte Carlo analysis either from anomalies in the Monte Carlo energy gradient $\Delta E/\Delta T$, or by variation in anisotropy. The I/III, I/V and II/III transitions are examples of the former; III/IV and IV/V are examples of the latter. Despite the differences in diffuse scattering patterns (and pore-network characteristics) throughout phase I, the entire phase field can be navigated without any discontinuity in energy or its derivative, or in anisotropy. Phase II is actually a physical mixture of separate components with $1/4$ and $1/2$ vacancy fractions: one has the Prussian blue structure, and the other is layered with tetragonal symmetry. Given this admixture, we do not expect the phase to be relevant to PBA chemistry. Phase V is also tetragonal, but is heavily disordered and may well be relevant to PBAs. Phase IV represents a competing mixture of (isotropic) III and (anisotropic) V components, and includes a morphotropic phase boundary⁴⁰. Phase VI

tetrahedral networks are shaded green; all others are collated for a given coordination number and shaded in grey. Coordination geometries are given at the top right: empty and filled circles denote occupied and vacant M' sites, respectively, and bold lines show connecting channels. Note the general preference for 90° pore angles at low J' and 120° angles at high J' .

contains a tetragonally ordered array of zig-zag pores. Our confidence in the detail of the phase diagram between phases I and VI is reduced by the difficulty of Monte Carlo equilibration at such low T' values.

In Fig. 3d we show a range of physical quantities calculated from our Monte Carlo configurations as a function of J' and T' . Some of these—for example, the Monte Carlo energy gradient $\log(\Delta E/\Delta T)$ or the diffuse scattering localization $L = \log[\Sigma(I^2)/(\Sigma I)^2]$ —serve primarily to highlight the phase boundaries, but others are particularly relevant to the transport properties of the PBA phases. For example, the tortuosity τ is a measure of the curvature of the internal pore space⁴¹. It affects the rate of mass transport (that is, the conductance)⁴²:

$$C \propto \frac{\rho}{\tau^2} \quad (2)$$

where ρ is the number of vacancy neighbour-pairs per formula unit⁴³. We find that C varies by a factor of two within the high-temperature disordered phase I and by yet another factor of two upon progression into lower-temperature phases. Even accessible pore volumes vary substantially—we calculate differences of greater than 25% for this same family of configurations. Moreover, in the anisotropic phases II, V and VI, transport depends on orientation.

This unexpected variability in micropore characteristics helps to explain the irreproducibility and diversity of sorption and storage properties observed experimentally. But it also highlights the opportunity for property optimisation via synthetic control over vacancy correlations—that is, defect engineering⁴⁴. For example, the value of

C should be maximized for battery materials, which might be achieved by combining low J' (for example, $M = \text{Cu}^{2+}$) and high T' (rapid precipitation). This analysis is consistent with the empirical identification of polycrystalline $\text{Cu}[\text{Fe}]$ (CuHCF) as a high-performance battery material¹⁰.

Our results identify many future challenges. We have focused on single-crystal samples because of the relative insensitivity of PXRD to the vacancy polymorphism of this family. So establishing a robust link between vacancy correlations and, for example, ion-storage capacity in hexacyanoferrates will require innovative approaches to measuring and interpreting diffuse scattering from microcrystalline samples. Serial femtosecond crystallography⁴⁵ or electron diffraction⁴⁶ may help. With access to vacancy-network models, it is possible that prospective sensitivity of powder pair distribution function measurements to vacancy correlations⁴⁷ may now be exploited. Our analysis has also been intentionally simplistic: we have not needed to invoke the role of alkali-metal inclusion, nor have we considered M' chemistry or cooperative Jahn–Teller effects⁴⁸. Yet these additional degrees of freedom must allow further chemical control over pore network characteristics. An obvious opportunity is to extend the phase fields of Fig. 3a as a function of alkali cation concentration; we might anticipate simpler behaviour as the vacancy fraction reduces. There are many variables that might be exploited in tailoring PBA network structures—for example, concentration, pH, crystal growth rate and media, temperature, speciation, solubility, competing ions and chelation—and establishing rules that link these variables to vacancy polymorphs represents an enormous challenge. Vacancy-network polymorphism may affect magnetic order, spin-state transitions, orbital order and photophysics. Moreover, any mechanistic understanding of double-metal cyanide catalysis will require characterization of particle surface structure, which may be substantially more complex than previously anticipated. And, stepping back, we might ask whether a similar hidden polymorphism plays a role in other microporous phases, such as metal–organic frameworks⁴⁹ or zeolites⁵⁰.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1980-y>.

- Kaye, S. S. & Long, J. R. Hydrogen storage in the dehydrated Prussian blue analogues $M_2[\text{Co}(\text{CN})_6]_2$ ($M = \text{Mn}, \text{Fe}, \text{Co}, \text{Ni}, \text{Cu}, \text{Zn}$). *J. Am. Chem. Soc.* **127**, 6506–6507 (2005).
- Haas, P. A. A review of information on ferrocyanide solids for removal of cesium from solutions. *Sep. Sci. Technol.* **28**, 2479–2506 (1993).
- Ohkoshi, S.-I. et al. High proton conductivity in Prussian blue analogues and the interference effect by magnetic ordering. *J. Am. Chem. Soc.* **132**, 6620–6621 (2010).
- Ferlay, S., Mallah, T., Ouahès, R., Veillet, P. & Verdaguer, M. A room-temperature organometallic magnet based on Prussian blue. *Nature* **378**, 701–703 (1995).
- Maurin, I. et al. Evidence for complex multistability in photomagnetic cobalt hexacyanoferrates from combined magnetic and synchrotron X-ray diffraction measurements. *Phys. Rev. B* **79**, 064420 (2009).
- Rykov, A. I., Li, X. & Wang, J. Crystal structure refinement of the electron-transfer-active potassium manganese hexacyanoferrates and isomorphous potassium manganese hexacyanocobaltates. *J. Solid State Chem.* **227**, 35–44 (2015).
- Bleuzen, A. et al. Photoinduced ferrimagnetic systems in Prussian blue analogues $C'_2\text{Co}_2[\text{Fe}(\text{CN})_6]_2$ ($C' = \text{alkali cation}$). 1. Conditions to observe the phenomenon. *J. Am. Chem. Soc.* **122**, 6648–6652 (2000).
- Le-Khac, B. & Chester, W. Highly active double metal cyanide catalysts. US patent 5,714,428 (1998).
- Peeters, A. et al. Zn–Co double metal cyanides as heterogeneous catalysts for hydroamination: a structure–activity relationship. *ACS Catal.* **3**, 597–607 (2013).
- Wessells, C. D., Huggins, R. A. & Cui, Y. Copper hexacyanoferrate battery electrodes with long cycle life and high power. *Nat. Commun.* **2**, 550 (2011).
- Pasta, M. et al. Full open-framework batteries for stationary energy storage. *Nat. Commun.* **5**, 3007 (2014).
- Ludi, A., Güdel, H.-U. & Rugg, M. The structural chemistry of Prussian blue analogs. A single-crystal study of manganese(II)hexacyanocobaltate(III), $\text{Mn}_2[\text{Co}(\text{CN})_6]_2 \cdot x\text{H}_2\text{O}$. *Inorg. Chem.* **9**, 2224–2227 (1970).
- Roque, J. et al. Porous hexacyanocobaltates(III): role of the metal on the framework properties. *Microporous Mesoporous Mater.* **103**, 57–71 (2007).
- Moritomo, Y., Igarashi, K., Kim, J. & Tanaka, H. Size-dependent cation channel in nanoporous Prussian blue lattice. *Appl. Phys. Express* **2**, 085001 (2009).
- Xiong, Q., Baychev, T. G. & Jivkov, A. P. Review of pore network modelling of porous media: experimental characteristics, network constructions and applications to reactive transport. *J. Contam. Hydrol.* **192**, 101–117 (2016).
- Grandjean, F., Samain, L. & Long, G. J. Characterization and utilization of Prussian blue and its pigments. *Dalton Trans.* **45**, 18018–18044 (2016).
- Keggin, J. F. & Miles, F. D. Structures and formulae of the Prussian blues and related compounds. *Nature* **137**, 577–578 (1936).
- Buser, H. J., Schwarzenbach, D., Petter, W. & Ludi, A. The crystal structure of Prussian blue: $\text{Fe}_4[\text{Fe}(\text{CN})_6]_3 \cdot x\text{H}_2\text{O}$. *Inorg. Chem.* **16**, 2704–2710 (1977).
- Herrn, F., Fischer, P., Ludi, A. & Hälg, W. Neutron diffraction study of Prussian blue, $\text{Fe}_4[\text{Fe}(\text{CN})_6]_3 \cdot x\text{H}_2\text{O}$. Location of water molecules and long-range magnetic order. *Inorg. Chem.* **19**, 956–959 (1980).
- Sharpe, A. G. *The Chemistry of Cyano Complexes of the Transition Metals* (Academic Press, 1976).
- Buser, H.-J., Ron, G., Ludi, A. & Engel, P. Crystal structure of cadmium hexacyanopalladate(IV). *J. Chem. Soc. Dalton Trans.* **23**, 2473–2474 (1974).
- Frisch, J. L. Notitia coerulei Berolinensis nuper inventi. *Misc. Berolin.* **1**, 377–378 (1710).
- Chadwick, B. M. & Sharpe, A. G. Transition metal cyanides and their complexes. *Adv. Inorg. Chem. Radiochem.* **8**, 83–176 (1966).
- Krap, C. P., Balmaseda, J., del Castillo, L. F., Zamora, B. & Reguera, E. Hydrogen storage in Prussian blue analogues: H_2 interaction with the metal found at the cavity surface. *Energy Fuels* **24**, 581–589 (2010).
- van der Marck, S. C. Calculation of percolation thresholds in high dimensions for FCC, BCC and diamond lattices. *Int. J. Mod. Phys. C* **9**, 529–540 (1998).
- Kaye, S. S. & Long, J. R. The role of vacancies in the hydrogen storage properties of Prussian blue analogues. *Catal. Today* **120**, 311–316 (2007).
- Roque-Malherbe, R., Lugo, F. & Polanco, R. Synthesis, structural elucidation and carbon dioxide adsorption on Zn(II) hexacyanoferrate(II) Prussian blue analogue. *Appl. Surf. Sci.* **385**, 360–367 (2016).
- Flambard, A., Köhler, F. H. & Lescouëzec, R. Revisiting Prussian blue analogues with solid-state MAS NMR spectroscopy: spin density and local structure in $[\text{Cd}_3[\text{Fe}(\text{CN})_6]_2] \cdot 15\text{H}_2\text{O}$. *Angew. Chem. Int. Ed.* **48**, 1673–1676 (2009).
- Calderon, H. A. & Reguera, E. HREM of porous Prussian blue type materials for hydrogen storage. *Microsc. Microanal.* **18**, 1468–1469 (2012).
- Franz, P. et al. Crystalline, mixed-valence manganese analogue of Prussian blue: magnetic, spectroscopic, X-ray and neutron diffraction studies. *J. Am. Chem. Soc.* **126**, 16472–16477 (2004).
- Chernyshov, D. & Bosak, A. Diffuse scattering and correlated disorder in manganese analogue of Prussian blue. *Phase Transit.* **83**, 115–122 (2010).
- Welberry, T. R. & Weber, T. One hundred years of diffuse scattering. *Crystallogr. Rev.* **22**, 2–78 (2016).
- Weber, T. & Simonov, A. The three-dimensional pair distribution function analysis of disordered single crystals: basic concepts. *Z. Krist.* **227**, 238–247 (2012).
- Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (Cornell Univ. Press, 1960).
- Paściak, M., Welberry, T. R., Kulda, J., Kempa, M. & Hlinka, J. Polar nanoregions and diffuse scattering in the relaxor ferroelectric $\text{PbMg}_{1/3}\text{Nb}_{2/3}\text{O}_3$. *Phys. Rev. B* **85**, 224109 (2012).
- Battle, P. D., Evers, S. I., Hunter, E. C. & Westwood, M. $\text{La}_2\text{Ni}_2\text{SbO}_6$: a relaxor ferromagnet. *Inorg. Chem.* **52**, 6648–6653 (2013).
- Zhdanov, H. Crystalline structure of $\text{Zn}(\text{CN})_2$. *C. R. Acad. Sci. URSS* **31**, 352–354 (1941).
- Shugam, E. & Zhdanov, H. The crystal structure of cyanides. The structure of $\text{Cd}(\text{CN})_2$. *Acta Physicochim. URSS* **20**, 247–252 (1945).
- Siebert, H. & Jentsch, W. Rhomboedrisch kristallisierende Zink-hexacyanometallate(III) $\text{Zn}_2[\text{M}(\text{CN})_6]_2$. *Z. Naturforsch. C* **36**, 123–124 (1981).
- Jaffe, B., Roth, R. S. & Marzullo, S. Piezoelectric properties of lead zirconate-lead titanate solid-solution ceramics. *J. Appl. Phys.* **25**, 809–810 (1954).
- Bear, J. *Dynamics of Fluids in Porous Media* (Dover, 1988).
- van der Linden, J. H., Narsillo, G. A. & Tordesillas, A. Machine learning framework for analysis of transport through complex networks in porous, granular media: a focus on permeability. *Phys. Rev. E* **94**, 022904 (2016).
- Epstein, N. On tortuosity and the tortuosity factor in flow and diffusion through porous media. *Chem. Eng. Sci.* **44**, 777–779 (1989).
- Fang, Z., Bueken, B., De Vos, D. E. & Fischer, R. A. Defect-engineered metal–organic frameworks. *Angew. Chem. Int. Ed.* **54**, 7234–7254 (2015).
- Zhang, T. et al. SFX analysis of non-biological polycrystalline samples. *IUCrJ* **2**, 322–326 (2015).
- Yun, Y., Zou, X., Hovmöller, S. & Wan, W. Three-dimensional electron diffraction as a complementary technique to powder X-ray diffraction for phase identification and structure solution of powders. *IUCrJ* **2**, 267–282 (2015).
- Chapman, K. W., Chupas, P. J., Maxey, E. R. & Richardson, J. W. Direct observation of adsorbed H_2 -framework interactions in the Prussian blue analogue $\text{Mn}^{II}_2[\text{Co}^{III}(\text{CN})_6]_2$: the relative importance of accessible coordination sites and van der Waals interactions. *Chem. Commun.* 4013–4015 (2006).
- Ojwang, D. O. et al. Structure characterization and properties of K-containing copper hexacyanoferrate. *Inorg. Chem.* **55**, 5924–5934 (2016).
- Liu, L. et al. Imaging defects and their evolution in a metal–organic framework at sub-unit-cell resolution. *Nat. Chem.* **11**, 622–628 (2019).
- Brunklaus, G., Koller, H. & Zones, S. I. Defect models of as-made high-silica zeolites: clusters of hydrogen bonds and their interaction with the organic structure-directing agents determined from ^1H double and triple quantum NMR spectroscopy. *Angew. Chem. Int. Ed.* **55**, 14459–14463 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Single-crystal growth

Single-crystal PBA samples were in all but one case grown using slow-diffusion methodologies. The exception was the Mn[Mn] crystal, for which we used the same sample reported in ref. ³¹. All other samples were prepared by counterdiffusion of aqueous solutions of a potassium hexacyanometallate(III) and a divalent transition-metal nitrate, chloride, sulphate or acetate. Precursor salts were used as supplied and without further purification; the identity and quantity of these salts are summarized in Extended Data Table 1. Crystals of Cd[Co], Mn[Co], Mn[Fe] and Zn[Co] were grown in H-cells, and those of Mn[Co], Cu[Co], and Co[Co] were grown from silica gel. For H-cell diffusion reactions, aqueous solutions of precursor salts (0.5 ml) were placed in the opposite arms of a glass H-cell. The cell was then filled with water, taking care not to disturb the solution–water interface. The H-cell was sealed and left undisturbed. Crystal growth was typically completed within 1–3 weeks. For gel-diffusion reactions, a transition-metal-impregnated gel was first prepared. Aqueous solutions of Na₂SiO₃ (2 M, 2.5 ml), M²⁺ salt (0.01–0.15 M, 2.5 ml), and acetic acid (2 M, 5 ml) were combined in a 15-ml centrifuge tube. The gel was allowed to set overnight. An approximately stoichiometric quantity of K₃[M'(CN)₆] was then dissolved in water (2.5 ml), and the corresponding solution layered carefully above the gel. The tube was sealed and the system left undisturbed. Crystal growth was typically completed within 3–7 d, with the first crystals appearing within 24 h. In all cases, care was taken not to dehydrate our samples.

Single-crystal diffuse scattering

Single-crystal diffuse scattering measurements were carried out using the I19 beamline at the Diamond Light Source (UK) and the BM01 beamline at the European Synchrotron Radiation Facility (France). Both beamlines are equipped with Pilatus 2M area pixel-counting detectors. The same data-acquisition strategy was used at both beamlines and consisted of a single 360° rotation scan around the omega axis. The key experimental parameters are summarized in Extended Data Table 2.

Determination of the crystal orientation and indexing and integration of the Bragg peaks were carried out using the package XDS⁵¹. Reconstruction of three-dimensional diffuse scattering was performed using the program Meerkat⁵². Air scattering was measured on an empty instrument, reconstructed in the same way as the signal and then subtracted from the signal. The diffuse scattering data so obtained were then averaged in the *m* $\bar{3}$ *m* Laue group. The reconstruction of the Mn[Mn] dataset from ref. ³¹ followed exactly the same procedure, using the original data frames.

Preparation of diffuse scattering inset images

The single-crystal X-ray diffuse scattering patterns shown in the insets of Fig. 2 were prepared from the three-dimensional scattering reconstruction using the projection method of ref. ⁵³. First, the *hk*0 section of each dataset was extracted, then the Bragg peaks were cut away, along with all surrounding thermal diffuse scattering. Next, the diffuse scattering was projected onto a single Brillouin zone. For this projection, square diffuse scattering patches with the diagonal corners (*h*, *k*) and (*h* + 2, *k* + 2) with *h* = 2*n*, *k* = 2*m* were taken and summed together. In the case of binary disorder, this procedure enables the removal of the contribution of the molecular form factor from the diffuse scattering. The resulting projection contains information only regarding the distribution of defects and does not include information about the chemical composition of those defects⁵³. This process is what enables direct comparison to the simulated diffuse scattering patches calculated from the Monte Carlo configurations.

3D- Δ PDF analysis

Diffuse scattering was analysed using the 3D- Δ PDF method^{33,54}. The experimental diffuse scattering was reconstructed as stated above, the background air scattering subtracted by using an empty instrument run, and an optimal scale coefficient selected manually. The resulting diffuse scattering was averaged in the *m* $\bar{3}$ *m* Laue group using outlier rejection as described by Blessing⁵⁵. Bragg peaks were removed using the ‘punch and fill’ procedure⁵⁶: spheres of intensity around the Bragg peaks were removed to ensure omission of thermal diffuse scattering contributions from subsequent analysis. The resulting holes were filled with the median intensity from a small surrounding region of reciprocal space. The crystals showed a large amount of thermal diffuse scattering around the Bragg peaks, and so the radius of spheres for punching and filling was chosen to be relatively large—approximately 0.5 reciprocal lattice units. Finally, the 3D- Δ PDF map was calculated using a fast Fourier transform. Quantitative 3D- Δ PDF refinement was carried out using the program Yell⁵⁷.

Here we give details for the representative case of the Co[Co] sample. The diffuse scattering map is presented in Extended Data Fig. 2a. Note that, owing to over-correction of the background, some pixels show negative intensities (marked red).

The 3D- Δ PDF map is presented in Extended Data Fig. 2b. The 3D- Δ PDF map gives the difference between the crystal pair distribution function and its Patterson function. This map should be interpreted in an analogous manner to the Patterson map; in particular, the signal at a position *uvw* corresponds to all the pairs of atoms in the structure which are separated by the vector components $u = x_i - x_j$, $v = y_i - y_j$, $w = z_i - z_j$, where *i* and *j* index atom pairs. The 3D- Δ PDF consists of positive and negative signals. Positive signals mean that corresponding interatomic pairs are present more often in the real structure than in the average structure; negative signals mean the opposite. For a more detailed introduction to the 3D- Δ PDF see ref. ³³.

Interpretation of the 3D- Δ PDF map in the current case is simplified by the presence of the strongly scattering atom cobalt in the partially vacant [Co(CN)₆]³⁻ moiety. The contribution of each interatomic pair is proportional to the product of the number of electrons in both atoms, and so the 3D- Δ PDF will be dominated by signals from pairs containing cobalt atoms. All atoms that are more likely to appear together with cobalt will give a positive contribution, and all of the atoms that tend to appear less often will be seen as negative signals.

The centre of the 3D- Δ PDF space is positive and represents all interatomic vectors within the [Co(CN)₆]³⁻ group. In the *uw*0 section, the signal appears as a cross. This is because the heavy cobalt atom is in the centre, and the four equatorial CN⁻ groups are around it. Similar crosses are located at positions corresponding to the face-centred lattice vectors. They correspond to the correlation between simultaneously present [Co(CN)₆]³⁻ groups. By contrast, the nearest neighbour at 0.5, 0.5, 0 shows a negative correlation, meaning that the probability of finding two [Co(CN)₆]³⁻ groups separated by this vector is less than 4/9 (the fraction observed in a completely random distribution of vacancies).

The program Yell enables refinement of the 3D- Δ PDF in terms of pair correlations. In order to speed up the refinement, the voxel size of diffuse scattering was increased in reciprocal space by binning sets of 5 × 5 × 5 voxels together. The final voxel resolution was 1/6 reciprocal lattice units, corresponding to the pair distribution function map containing correlations spanning the nearest three unit cells in the *x*, *y* and *z* directions.

In our modelling of the 3D- Δ PDF, we have assumed that two-thirds of the M' sites are occupied by [Co(CN)₆]³⁻ ions and the remaining one-third are occupied by an H₂O-filled vacancy (Extended Data Fig. 3). The latter was modelled with six structural water molecules and four zeolitic water molecules. The model 3D- Δ PDF is shown in the right-hand panel of Extended Data Fig. 2b. It is clear that the model accounts well for

the majority of the observed features. The resulting correlation coefficients for the first 44 neighbours are listed in Extended Data Table 3.

Monte Carlo simulations and analysis

Monte Carlo simulations were carried out using a parallel tempering approach⁵⁸ implemented within custom-written code. We note for context the use elsewhere of Monte Carlo methods for interpreting single-crystal diffuse scattering^{32,59}. In our case, for each J' value, an ensemble of 129 configurations was generated and Monte Carlo simulations were carried out at a suitable distribution of temperatures T' ($0.75 < T' < 4.97329$, with $T'_i = 1.014889 \times T'_{i-1}$; that is, evenly spread in $\log T'$). Each configuration represented a $12 \times 12 \times 12$ supercell of the fcc unit cell, containing a total of 6,912 sites. Configurations were initialized with a random distribution of vacancies, such that exactly one-third of the sites were vacant. The Monte Carlo steps involved swap moves: two sites, one occupied and one vacant, were selected at random and their contents swapped. In addition to these Monte Carlo steps, the algorithm involved replica exchange. An attempt for one replica exchange was performed every four Monte Carlo steps. For this, two reservoirs with nearest temperatures were selected at random, and the temperature swap was performed with probability $p = \exp[(E_1 - E_2)/(T_1^{-1} - T_2^{-1})]$. The configurations were equilibrated for 100 epochs (one epoch corresponds to the number of steps required to visit each site twice on average), following production steps of 80 epochs each. The thermodynamic quantities (E , E^2) were sampled at each Monte Carlo step, and all other quantities (tortuosity, diffuse scattering, etc.) were calculated from 40 configurations separated from one another by two epochs. Convergence was determined by the convergence of the Monte Carlo energies of the models. It was found that almost all configurations converged, except for a few configurations of polymorphs II, IV and VI at the very lowest sampled temperatures. However, the diffuse scattering from unconverged configurations of phase IV showed sharp streaks parallel to the reciprocal axes \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* that were very similar to the streaks observed experimentally in Mn[Co]. In our view, the experimental crystals are also not necessarily all at thermodynamic equilibrium, and so we decided to keep the simulation without change. The diffuse scattering patterns shown in Fig. 3 were calculated using a fast Fourier transform, averaged in the $m\bar{3}m$ Laue group. Only the hkl planes were extracted and shown.

Surface area and accessible pore volume calculations were calculated using the Zeo++ code⁶⁰ for small configurations, and a related custom-written code for larger configurations.

Tortuosity was calculated as the average of the distance from a vacancy site to a plane along the percolation channel divided by the 'flight' distance to the plane: $\tau = d_i/d_i^a$. Here, d_i is the length along the percolating links from the vacancy to a plane and d_i^a is the 'flight' distance of the same vacancy to the plane. The values of tortuosity were calculated in each of the $\langle 100 \rangle$ directions and then averaged. Because vacancies are connected with each other in the $\langle 110 \rangle$ directions, the smallest tortuosity achievable in this structure is $\sqrt{2}$.

Strain effects mean the energy of charged defects may not always scale as the square of the charge (see, for instance, ref. ⁶¹), and so we have checked the robustness of our results with the modified Hamiltonian

$$E = \sum_{\mathbf{r} \in \{M\}} \left[J_1 \left| 4 - \sum_{\mathbf{r}' \in \frac{1}{2}\langle 100 \rangle} e_{\mathbf{r}+\mathbf{r}'} \right| + \frac{J_2}{2} \sum_{\mathbf{r}' \in \frac{1}{2}\langle 100 \rangle} (e_{\mathbf{r}+\mathbf{r}'} - e_{\mathbf{r}-\mathbf{r}'})^2 \right] \quad (2)$$

where, compared to equation (1), we have used the absolute value in the J_1 term. The resulting phase diagram is essentially the same, albeit the simulations converge considerably more slowly, owing to additional frustration in phase V. The diffuse scattering from this model is shown in Extended Data Fig. 4.

Surface area and accessible pore volume calculations were performed using the Zeo++ code⁶¹. We found the code to be prohibitively

slow for the large $12 \times 12 \times 12$ configurations, and so we used smaller configurations ($6 \times 6 \times 6$) for initial calculations. In these calculations we observed that both the accessible volume and the accessible surface areas were directly proportional to the number of accessible vacancies in the structure. Consequently, for the larger $12 \times 12 \times 12$ configurations, we obtained the final values using the following relations: $S_a = 1.551 \times N_a/N_t$ [$\text{m}^2 \text{g}^{-1}$] and $V_a = 0.074 \times N_a/N_t$ [ml g^{-1}]. Here, V_a is the total accessible volume, S_a is the accessible surface, N_a is the number of accessible vacancies and N_t is the total number of vacancies in the simulation box.

Location of projections within the phase diagram

For all but one PBA system, the corresponding experimental diffuse scattering 'tiles' shown in Fig. 3 were positioned to minimize the difference between experimental and model intensities. We determined these differences using the diffuse scattering R factor:

$$R = \frac{\sum_{hkl} \{I_e(hkl) - [sI_m(hkl) + c]\}^2}{\sum_{hkl} [I_e(hkl)]^2} \quad (3)$$

Here, s is the scale coefficient of the model, c is a constant background, and I_e and I_m are the experimental and model scattering intensities, respectively. The parameters s and c were determined by linear minimization of the R factor.

The exception to this approach was in the case of Mn[Co]. Here, our experimental data showed the presence of sharp streaks of diffuse scattering parallel to the \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* directions. The R -factor approach described above could not correctly place this tile, owing to the absence of accurate modelling of the experimental resolution function. Instead, this particular tile was placed within area IV of the phase diagram, which we felt best accounted for the qualitative features of the experimental data. The projected experimental diffuse scattering patterns and the closest matching tiles from the model phase diagram are compared in Extended Data Fig. 5.

As a final point we note that comparison metrics alternative to the one we propose here are easily envisaged. We tested a number of these metrics during the course of our own analysis and found that different approaches gave slightly different positions for the various diffuse scattering tiles. Nevertheless, in essentially all cases the general features noted in the main text were preserved; for example, that Zn[Co] and Cd[Co] were placed on the right-hand side of the phase diagram, Cu[Co] on the left, and the $M = \text{Mn, Co}$ samples were near $J' = 1$.

Additional diffraction features in PBAs

In addition to the vacancy-driven diffuse scattering and satellites described previously, we observed diffraction features that are not related to vacancy order. These intensities are not covered by our model (understandably), but we expand here on why their presence does not affect our analysis or the conclusions drawn.

For the Cu[Co] sample, we observed not only vacancy-driven diffuse scattering and Bragg peaks corresponding to P centring, but also additional Bragg peaks near the half-integer positions. The indices of such reflections are approximately equal to $(h, k, 0.9l + 0.5)$; that is, they come from a tetragonally distorted crystallite with unit-cell parameters $a' = a_{\text{main}}$, and $c' = 1.1a_{\text{main}}$, where a_{main} is the unit-cell parameter of the primary crystal phase (Extended Data Fig. 6). Other Bragg peaks from this sample showed no signs of such tetragonal distortion and the total intensity of additional reflections was less than 1% of the total main reflections, and so we can conclude that these additional reflections come from a tetragonally distorted impurity precipitate which is coherent with respect to the primary cubic matrix.

Some crystals also showed diffuse scattering centred at positions of the type $(h + 1/3, k + 1/3, l)$ (Extended Data Fig. 7). In particular, these additional signals were visible for the Cd[Co], Mn[Mn], Mn[Fe] and Mn[Co] crystals, but not for the Mn[Co], Zn[Co], Cu[Co] or Co[Co]

crystals. The intensity of this diffuse scattering was zero for $l = 0$ (in the planes passing through the centre of reciprocal space), and increased with the magnitude of l . This is a typical signature of diffuse scattering from displacive correlations, with displacement polarized along the c direction. Interestingly, in the $\text{Mn}[\text{Co}]'$ and $\text{Cd}[\text{Co}]$ samples the diffuse scattering intensities follow tetragonal symmetry, instead of the cubic symmetry of the Bragg peaks. Note that because displacive diffuse scattering is absent in the $hk0$, $h0l$ and $0kl$ sections it has no contribution to the projected diffuse scattering tiles from Figs. 2 and 3 and thus does not influence the analysis of vacancy distributions.

Data availability

The raw data on which this manuscript is based are openly available for download from <https://doi.org/10.5287/bodleian:8JB5XgybE>. These data include the scattering images given in Fig. 2 and the Monte Carlo configurations from which Figs. 3 and 4 are derived.

Code availability

All custom code used in this study was developed using widely available algorithms. Copies of the code used can be obtained upon request.

51. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
52. Logvinovich, D. & Simonov, A. Meerkat. <https://github.com/aglie/meerkat> (2015).
53. Schmidt, E. & Neder, R. B. Diffuse single-crystal scattering corrected for molecular form factor effects. *Acta Crystallogr. A* **73**, 231–237 (2017).
54. Simonov, A., Weber, T. & Steurer, W. Experimental uncertainties of three-dimensional pair distribution function investigations exemplified on the diffuse scattering from a tris-tert-butyl-1,3,5-benzene tricarboxamide single crystal. *J. Appl. Crystallogr.* **47**, 2011–2018 (2014).
55. Blessing, R. H. Outlier treatment in data merging. *J. Appl. Crystallogr.* **30**, 421–426 (1997).
56. Kobas, M., Weber, T. & Steurer, W. Structural disorder in the decagonal Al–Co–Ni. I. Patterson analysis of diffuse X-ray scattering data. *Phys. Rev. B* **71**, 224205 (2005).

57. Simonov, A., Weber, T. & Steurer, W. Yell: a computer program for diffuse scattering analysis via three-dimensional delta pair distribution function refinement. *J. Appl. Crystallogr.* **47**, 1146–1152 (2014).
58. Earl, D. J. & Deem, M. W. Parallel tempering: theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **7**, 3910–3916 (2005).
59. Welberry, T. R., Goossens, D. J., Edwards, A. J. & David, W. I. F. Diffuse X-ray scattering from benzil, $\text{C}_{18}\text{H}_{10}\text{O}_2$: analysis via automatic refinement of a Monte Carlo model. *Acta Crystallogr. A* **57**, 101–109 (2001).
60. Willems, T. F., Rycroft, C. H., Kazi, M., Meza, J. C. & Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **149**, 134–141 (2012).
61. Aschauer, U. & Spaldin, N. A. Interplay between strain, defect charge state, and functionality in complex oxides. *Appl. Phys. Lett.* **109**, 031901 (2016).

Acknowledgements A.S. and A.L.G. gratefully acknowledge financial support from the Leverhulme Trust UK (grant no. RPG-2015-292), and T.D.B. acknowledges F.W.O.–Vlaanderen (Research Foundation Flanders) for a Postdoctoral Fellowship. A.S. thanks the Swiss National Science Foundation for Ambizione and PostDoc Mobility Fellowships (PZ00P2_180035, P2EZP2_155608) and Diamond Light Source for the provision of beamtime (MT13639, MT20876, CY22610). M.L.R.G. thanks the Consejo Nacional de Ciencia y Tecnología (Mexico) for a scholarship. A.L.G. thanks the European Research Council for funding (grant nos 279705 and 788144), P. D. Battle (Oxford) and A. R. Overy (Oxford) for valuable discussions, N. P. Funnell (ISIS), J. A. Hill (UCL) and C. S. Coates (Oxford) for assistance with single-crystal growth, and A. L. Thompson for assistance with synchrotron measurements.

Author contributions A.S., T.D.B., H.-B.B. and A.L.G. designed the research. H.L.B.B., M.L.R.G., H.J.G. and T.D.B. synthesized the materials. A.S., D.C., A.B. and H.-B.B. measured the single-crystal diffuse scattering patterns. A.S. performed the 3D- Δ PDF refinement. A.S. and A.L.G. developed and implemented the Monte Carlo model. A.S. and T.D.B. calculated percolation properties. A.S. and A.L.G. wrote the manuscript, with input from all authors.

Competing interests The authors declare no competing interests.

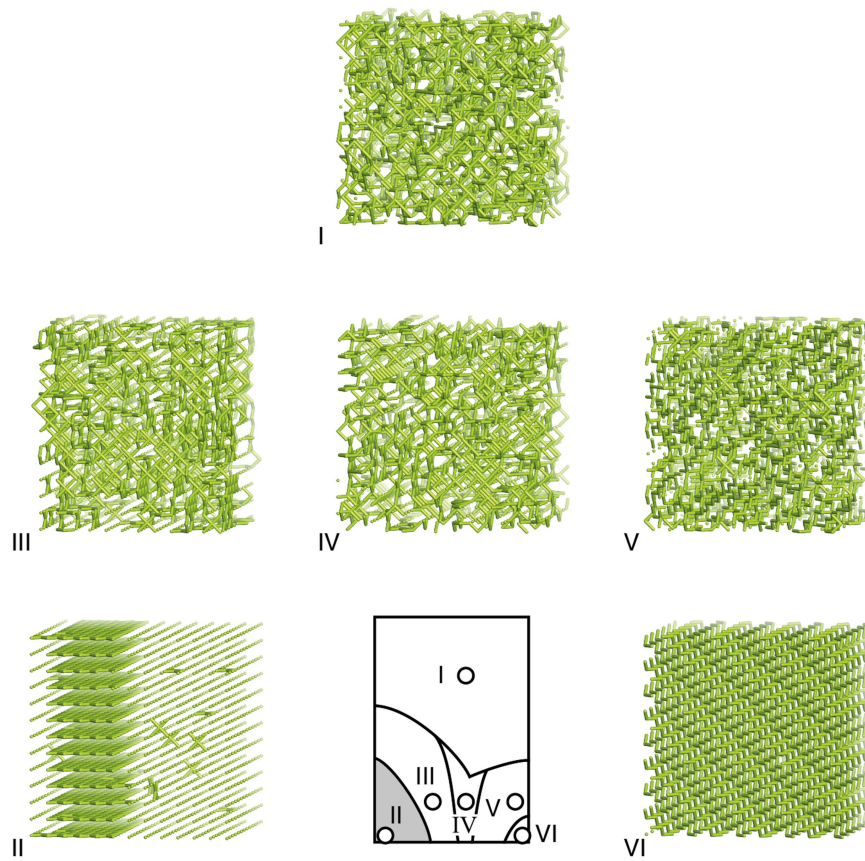
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1980-y>.

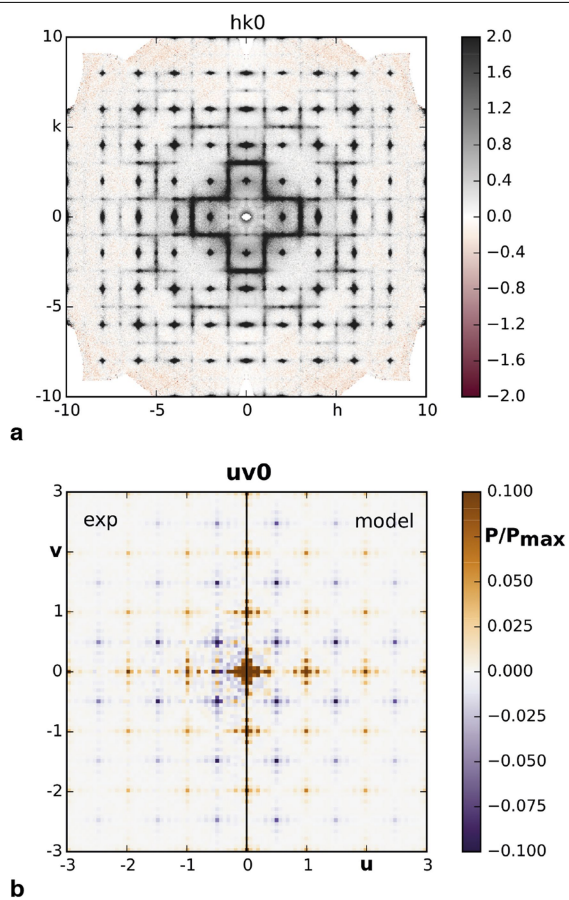
Correspondence and requests for materials should be addressed to A.L.G.

Peer review information *Nature* thanks Simon Billinge, T. Richard Welberry and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

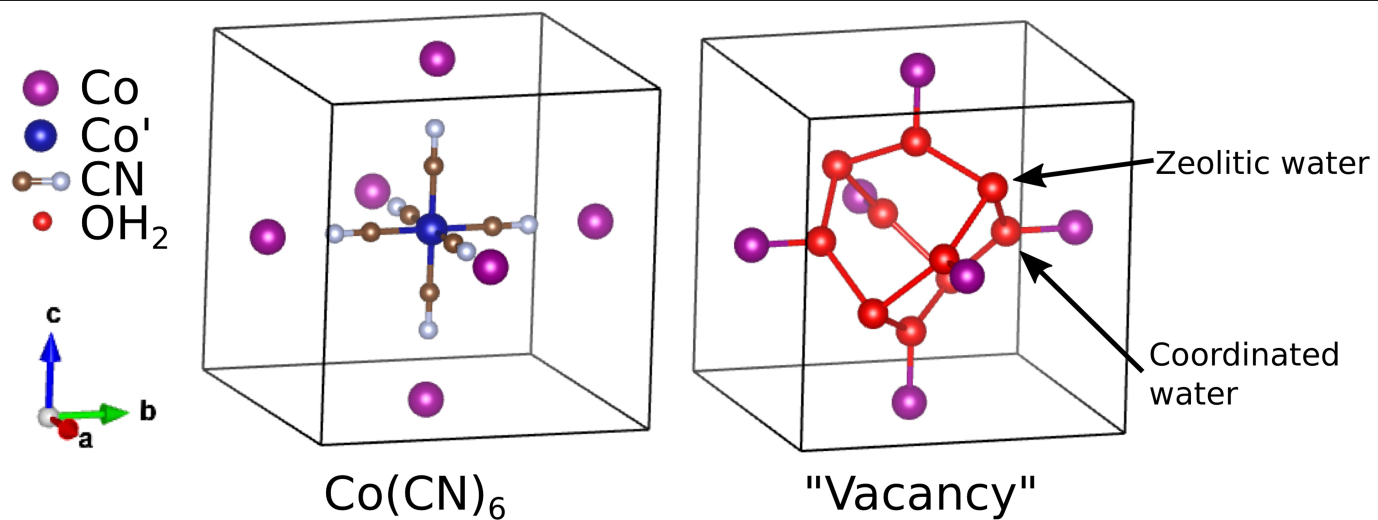
Reprints and permissions information is available at <http://www.nature.com/reprints>.



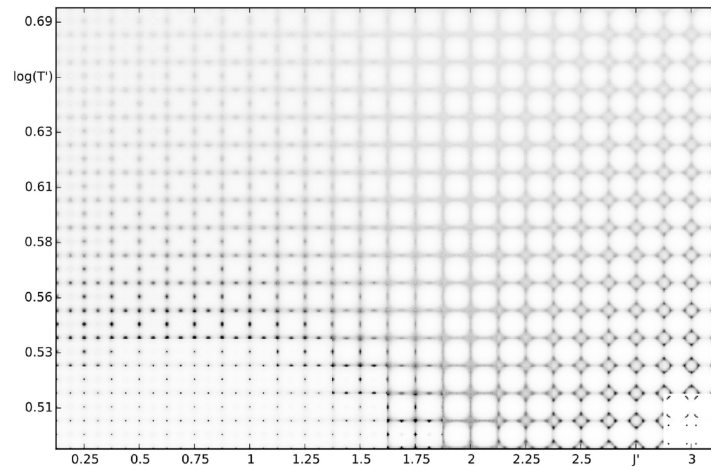
Extended Data Fig. 1 | Representative pore networks. Representative pore networks for each phase within the Monte Carlo simulated phase diagram.



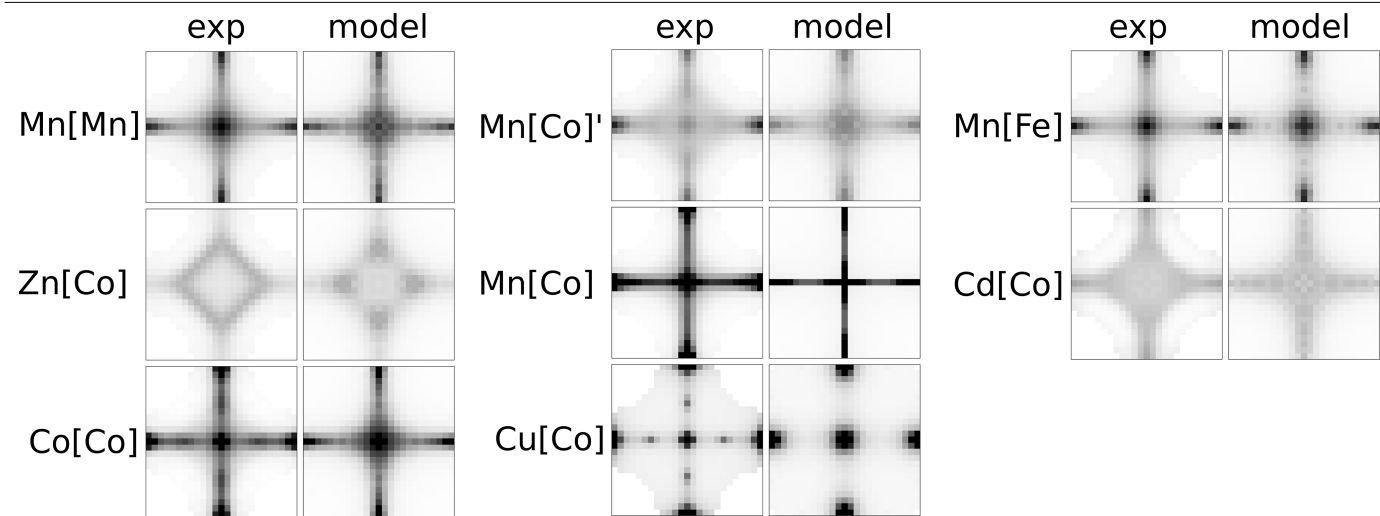
Extended Data Fig. 2 | Representative 3D- Δ PDF. **a**, Experimental diffuse scattering from the Co[Co] sample, $hk0$ section; voxel size is 1/30 reciprocal lattice units. **b**, Experimental and model 3D- Δ PDF map of the Co[Co] sample, $uv0$ section.



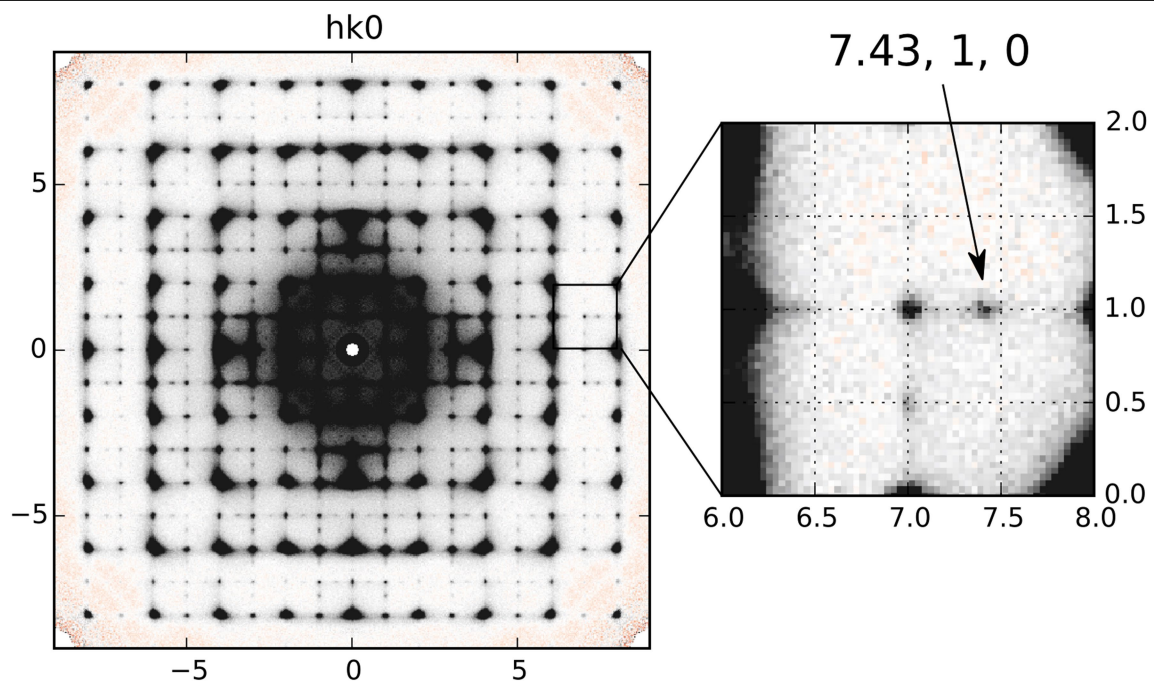
Extended Data Fig. 3 | M'-site models. The structure of the [Co(CN)₆]³⁻ and 'vacancy' moieties used in our Co[Co] model.



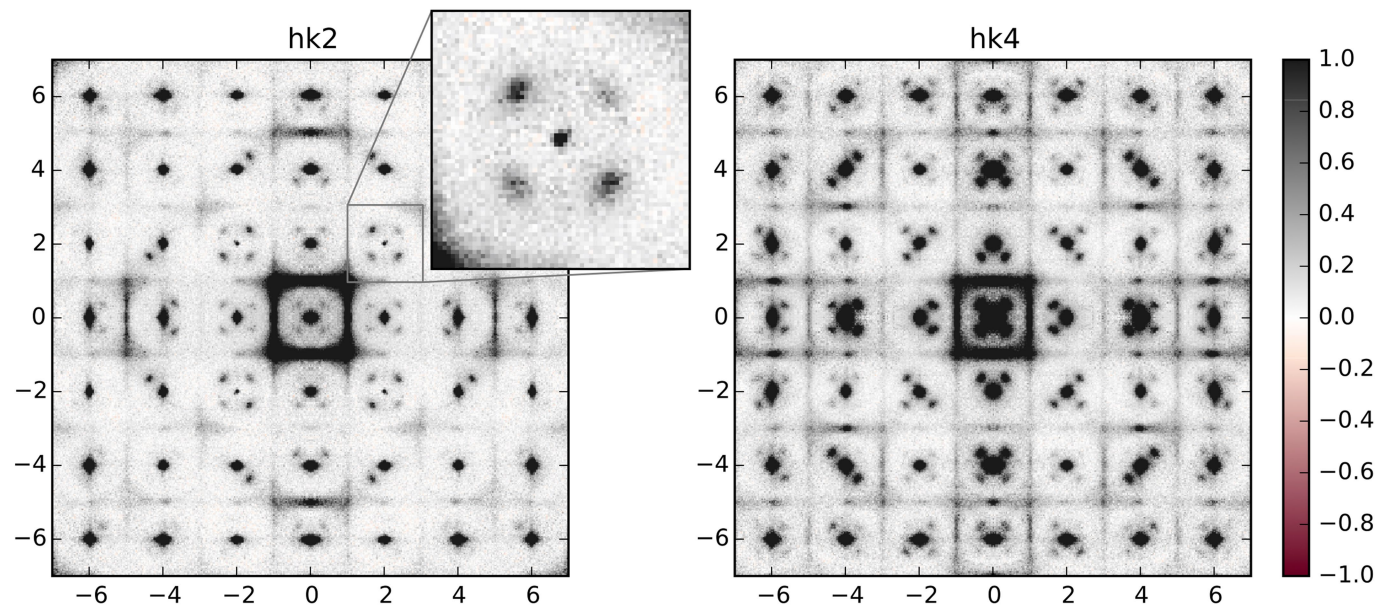
Extended Data Fig. 4 | Alternative diffuse scattering phase map. Diffuse scattering calculated with the modified model Hamiltonian; equation (2).



Extended Data Fig. 5 | Diffuse scattering tiles. Comparison of projected experimental diffuse scattering with the model diffuse scattering tiles for our various PBA samples; compare with Fig. 3.



Extended Data Fig. 6 | Satellite reflections. Satellite reflections in our Cu[Co] sample. The inset shows one specific satellite at $(7.43, 1, 0)$.



Extended Data Fig. 7 | Inelastic scattering. Diffuse scattering in the Mn[Co]' sample. Note that the intensity of the diffuse scattering at the $(h+1/3, k+1/3, l)$ positions increases with increasing l ; the scattering in the $hk4$ layer (right) is stronger than in the $hk2$ layer (left).

Extended Data Table 1 | Synthesis summary

PBA	Method	M ²⁺ salt	<i>n</i> (M) (mmol)	K ₃ [M'(CN) ₆]	<i>n</i> (M') (mmol)
Cu[Co]	Gel	CuSO ₄ ·5H ₂ O	0.03	K ₃ [Co(CN) ₆]	0.02
Co[Co]	Gel	CoSO ₄	0.30	K ₃ [Co(CN) ₆]	0.20
Mn[Co]	Gel	MnSO ₄ ·H ₂ O	0.12	K ₃ [Co(CN) ₆]	0.12
Mn[Co]'	H-Cell	MnSO ₄ ·H ₂ O	0.202	K ₃ [Co(CN) ₆]	0.134
Mn[Fe]	H-Cell	MnSO ₄ ·H ₂ O	0.131	K ₃ [Fe(CN) ₆]	0.131
Cd[Co]	H-Cell	Cd(CH ₃ COO) ₂	0.156	K ₃ [Co(CN) ₆]	0.104
Zn[Co]	H-Cell	Zn(NO ₃) ₂ ·6H ₂ O	0.10	K ₃ [Co(CN) ₆]	0.067

Precursor compositions and quantities used for PBA synthesis.

Extended Data Table 2 | Data collection strategies

PBA	Beamline	λ (Å)	Δr (Å)	Oscillation range (°)	$\Delta\phi$ (°)
Cu[Co]	I19	0.6889	0.83	360	0.1
Co[Co]	I19	0.6889	0.83	360	0.1
Mn[Co]	I19	0.6889	0.83	360	0.1
Mn[Co]'	I19	0.6889	0.83	360	0.1
Mn[Fe]	I19	0.6889	0.83	360	0.1
Cd[Co]	BM01	0.6975	0.64	360	0.1
Zn[Co]	I19	0.6889	0.83	360	0.1
Mn[Mn]	BM01 (Mar345)	0.71	1.3	180	0.5

Single-crystal X-ray diffuse scattering data collection strategies. Included in the last row are, for comparison, the relevant values for the Mn[Mn] sample reported in ref. ³¹. Δr , diffraction resolution; $\Delta\phi$, angular step size.

Extended Data Table 3 | 3D-ΔPDF model

<i>u</i>	<i>v</i>	<i>w</i>	<i>s</i>	<i>u</i>	<i>v</i>	<i>w</i>	<i>s</i>	<i>u</i>	<i>v</i>	<i>w</i>	<i>s</i>	<i>u</i>	<i>v</i>	<i>w</i>	<i>s</i>
0	0	0	1	0	1.5	2.5	−0.013	0.5	1.5	3	0.00065	1	3	3	0.002
0	0	1	0.15	0	2	2	0.019	0.5	2	2.5	−0.0025	1.5	1.5	2	−0.0034
0	0	2	0.071	0	2	3	0.014	0.5	2.5	3	−0.0012	1.5	1.5	3	−0.0012
0	0	3	0.047	0	2.5	2.5	−0.0097	1	1	1	−0.0057	1.5	2	2.5	−0.0025
0	0.5	0.5	−0.073	0	3	3	0.01	1	1	2	0.0033	1.5	2.5	3	−0.00025
0	0.5	1.5	−0.047	0.5	0.5	1	0.0056	1	1	3	0.0015	2	2	2	0.0051
0	0.5	2.5	−0.028	0.5	0.5	2	0.0078	1	1.5	1.5	0.0038	2	2	3	0.0047
0	1	1	0.033	0.5	0.5	3	0.0074	1	1.5	2.5	0.0002	2	2.5	2.5	−0.0012
0	1	2	0.025	0.5	1	1.5	−0.0025	1	2	2	0.0047	2	3	3	0.0033
0	1	3	0.018	0.5	1	2.5	−0.0043	1	2	3	0.0033	2.5	2.5	3	−0.0012
0	1.5	1.5	−0.021	0.5	1.5	2	−0.0012	1	2.5	2.5	−0.0007	3	3	3	−0.003

Refined 3D-ΔPDF correlation coefficients *s* of the first 44 neighbours.

Premature mortality related to United States cross-state air pollution

<https://doi.org/10.1038/s41586-020-1983-8>

Received: 1 December 2017

Accepted: 1 November 2019

Published online: 12 February 2020

Irene C. Dedoussi^{1,2}, Sebastian D. Eastham^{1,3}, Erwan Monier^{3,4} & Steven R. H. Barrett^{1,3*}

Outdoor air pollution adversely affects human health and is estimated to be responsible for five to ten per cent of the total annual premature mortality in the contiguous United States^{1–3}. Combustion emissions from a variety of sources, such as power generation or road traffic, make a large contribution to harmful air pollutants such as ozone and fine particulate matter (PM_{2.5})⁴. Efforts to mitigate air pollution have focused mainly on the relationship between local emission sources and local air quality². Air quality can also be affected by distant emission sources, however, including emissions from neighbouring federal states^{5,6}. This cross-state exchange of pollution poses additional regulatory challenges. Here we quantify the exchange of air pollution among the contiguous United States, and assess its impact on premature mortality that is linked to increased human exposure to PM_{2.5} and ozone from seven emission sectors for 2005 to 2018. On average, we find that 41 to 53 per cent of air-quality-related premature mortality resulting from a state's emissions occurs outside that state. We also find variations in the cross-state contributions of different emission sectors and chemical species to premature mortality, and changes in these variations over time. Emissions from electric power generation have the greatest cross-state impacts as a fraction of their total impacts, whereas commercial/residential emissions have the smallest. However, reductions in emissions from electric power generation since 2005 have meant that, by 2018, cross-state premature mortality associated with the commercial/residential sector was twice that associated with power generation. In terms of the chemical species emitted, nitrogen oxides and sulfur dioxide emissions caused the most cross-state premature deaths in 2005, but by 2018 primary PM_{2.5} emissions led to cross-state premature deaths equal to three times those associated with sulfur dioxide emissions. These reported shifts in emission sectors and emission species that contribute to premature mortality may help to guide improvements to air quality in the contiguous United States.

Long-term exposure to fine particulate matter (PM_{2.5}) and ozone leads to an increased risk of premature death^{7–12}. Indeed, PM_{2.5} and ozone are the most prominent known causes of early deaths associated with outdoor air pollution, resulting in more than 90% of total air-pollution-related mortalities^{8,11}. For this reason, PM_{2.5} and ozone have become the predominant pollutants for quantifying air quality². These pollutants form mainly through atmospheric chemical reactions following the release of precursor emissions. PM_{2.5}, which consists of particles and liquid droplets, forms from gaseous precursor emissions of nitrogen oxides (NO_x), sulfur oxides (SO_x), ammonia (NH₃), and others. PM_{2.5} can also be emitted directly, as in the case of black carbon. Ozone forms from gaseous precursor emissions of NO_x and volatile organic compounds (VOCs). The adverse health impacts due to exposure to PM_{2.5} and ozone can therefore be attributed to the precursor emissions that lead to their formation. Such attribution is useful, as it is these

emissions that can be directly controlled, rather than the exposure that results from them.

Combustion emissions constitute the largest source of anthropogenic emissions in the USA, and therefore contribute to the formation of PM_{2.5} and ozone². The health impacts attributable to these emissions have been estimated in various studies^{6,13,14}, with estimates varying between 90,000 and 360,000 early deaths per year. In the context of the Environmental Protection Agency (EPA) Cross-State Air Pollution Rule (CSAPR) and individual state regulation, measures to further reduce the health impacts of pollution would benefit from a greater understanding of which sectors and which states are responsible for the health impacts in every other state.

Prior studies have investigated parts of this problem. One study⁶ estimated the sources of US PM_{2.5} pollution impacts on a fine scale, with other work focusing on the roles of individual emission sectors¹⁵

¹Laboratory for Aviation and the Environment, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Section Aircraft Noise and Climate Effects, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands. ³Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Department of Land, Air and Water Resources, University of California, Davis, Davis, CA, USA. *e-mail: sbarrett@mit.edu

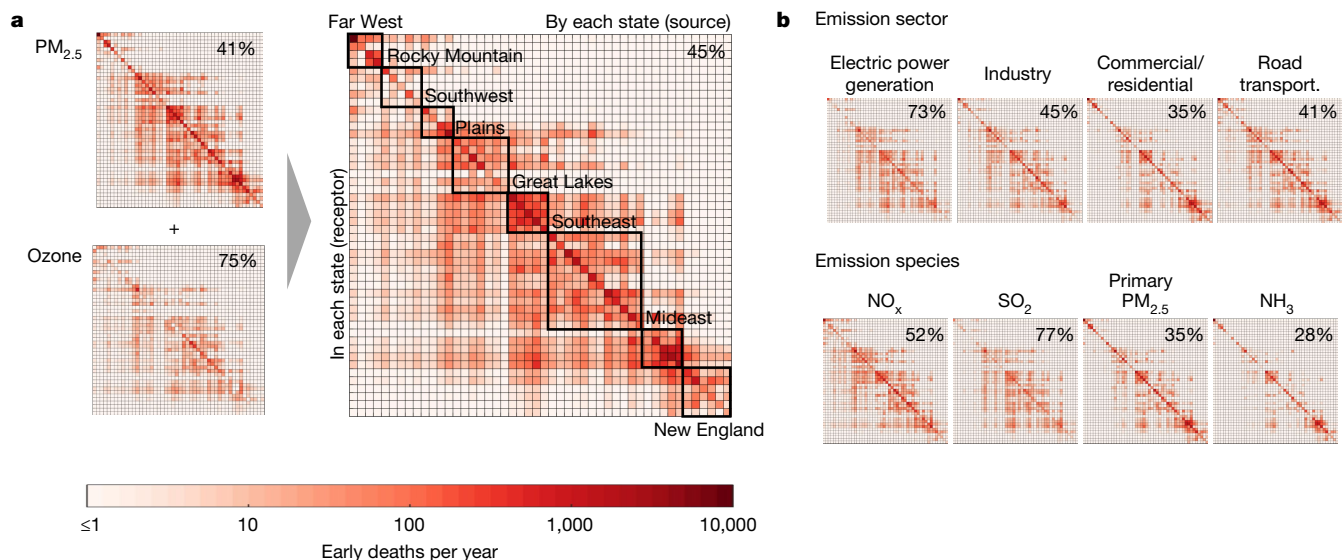


Fig. 1 | Early-death source–receptor matrices for 2011. **a**, Source–receptor matrix showing total early deaths per year for 48 × 48 states (right), and its breakdown into PM_{2.5}- and ozone-attributable impacts (left). **b**, Source–receptor early-death attribution to emission sectors (top) and emission species (bottom) that lead to the formation of PM_{2.5} and/or ozone. States are grouped into US Bureau of Economic Analysis regions²⁴ and ordered west (left) to east

(right) (ordering presented in Extended Data Fig. 1). Boxed percentages represent the fraction of impacts that occur out of the state that caused the corresponding emissions. Obtaining the summarized matrices shown using conventional approaches ('forward difference') would require 433-year-long simulations. Extended Data Fig. 2 presents corresponding matrices for 2005 and 2018.

or species^{16,17} for either pollutant. Numerous other studies have focused on examining the roles of different emission sectors^{13,18} or species^{19,20}, without quantifying the aspect of pollution exchange. Variations in time have also been discussed²¹. In all cases these studies have focused on only one or two of the dimensions of the problem (emission sector, emission species, pollutant and exchange), and no previous work has integrated these aspects together into a single study. As such, to date there has been no assessment of cross-state pollution exchange that quantifies the influence, by sector and chemical species, of each state on every other state's health risk, using detailed chemistry–transport modelling and including both PM_{2.5} and ozone.

In this work, we estimate the pollution exchange between the 48 contiguous US states, and form source–receptor relationships between them for combustion emissions from seven sectors: electric power generation; industry; commercial/residential; road transportation; marine; rail; and aviation. The commercial/residential sector includes residential combustion (for example, of biomass), nonindustrial commercial and institutional processes, and waste treatment, among other sources. This analysis yields estimates for the number of early deaths due to PM_{2.5} (primary and secondary, excluding secondary organic aerosols) and ozone exposure in every state, with attribution of impacts to each sector and each emitted chemical species from every state. We estimate combustion emissions for the seven sectors for 2005 (based on the 2005 National Emissions Inventory (NEI)), 2011 (based on NEI2011) and 2018 (based on the NEI2011 forecast), and present these findings in Extended Data Table 1. Lists of the specific sources that are grouped in each sector are included in the associated data repository (see Methods). The impacts of these emissions on each state's air quality are then quantified using receptor-oriented atmospheric sensitivities from the adjoint of the GEOS-Chem chemistry–transport model²² (see Methods).

We calculate the pollution exchange between every state pair for the contiguous US for every combination of emission sector, PM_{2.5} or ozone precursor emission species, and year. The 2011 source–receptor relations for the two pollutants and the total impacts are summarized in Fig. 1a. Matrices for different sectors and emission species are presented in Fig. 1b. Source–receptor matrices for all three years are presented in Extended Data Fig. 2.

The relative percentage of total impacts that occurred outside of the emitting state decreased with time, from 53% in 2005, to 45% in 2011 and 41% in 2018, meaning that there has been a declining relative magnitude of cross-state impacts. This fraction varies substantially between sectors. Electric power generation is the only sector that is regulated by the CSAPR, and has the highest out-of-state impacts as a fraction relative to in-state impacts: on average, approximately 70% of early deaths from this sector occur outside of the state that caused the emissions. However, with reductions in emissions from electric power generation, by 2018 there were 70% fewer out-of-state early deaths (approximately 13,000 fewer early deaths) by comparison with 2005. Road transportation, industry and commercial/residential emissions resulted in higher cross-state early deaths in 2018 than electric power generation (by 28%, 42% and 74% respectively), but are not regulated by the CSAPR at present. Although PM_{2.5} and ozone impacts can vary by +125% to −65% depending on the specific choice of concentration-response function (see Methods), this disagreement does not affect the net pollution exchange between states and the impacts attributable to each sector.

The results presented in Fig. 1a, b reflect both PM_{2.5}- and ozone-attributable early deaths. Although the number of early deaths per additional unit of emission is approximately eight times higher for PM_{2.5} than for ozone (not accounting for nonlinear interactions; see Methods), ozone impacts are typically transported farther. The fraction of PM_{2.5} impacts that happen out of the state that caused them was approximately 41% for 2011, compared with approximately 75% for ozone for the same year. The full source–receptor matrices for each sector–year and species–year combination are included in the data repository (see Methods).

The fact that the source–receptor matrices, presented in Fig. 1, are not symmetric about the diagonal implies that there is a net imbalance in the exchange of early deaths between the US states. Figure 2 presents this exchange in terms of the air-quality-related early deaths per capita because of emissions from each state (Fig. 2a) and occurring within each state (Fig. 2b), as well as the net exchange between states (Fig. 2c). A positive value in Fig. 2c indicates that a given state is a net 'exporter' of early deaths—that is, that emissions in that state cause

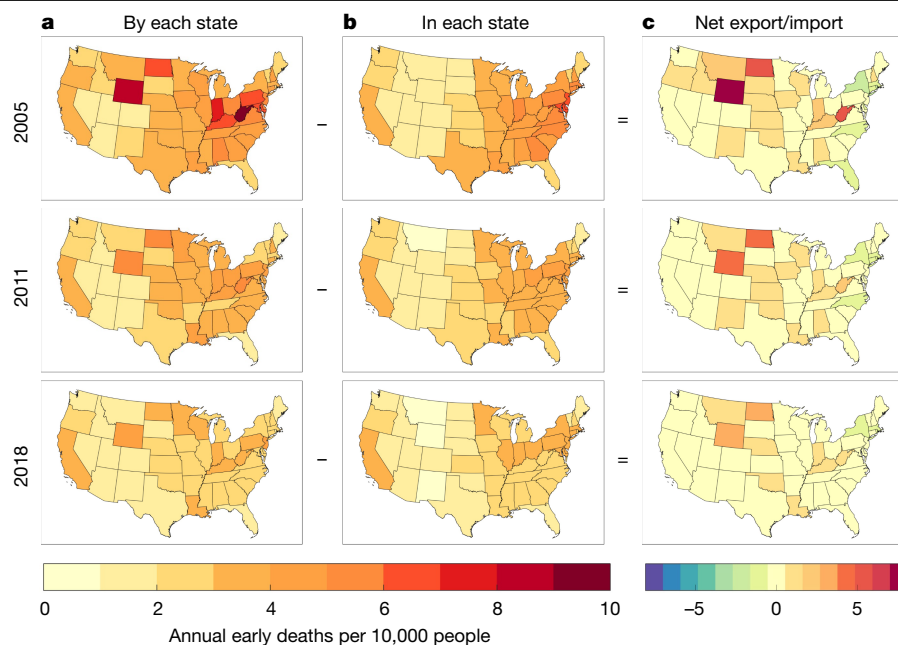


Fig. 2 | Total annual early deaths caused per 10,000 people for 2005, 2011 and 2018. The left plots (a) show the total aggregate early deaths caused by emissions in each state, divided by the population of the emitting state. The middle plots (b) show the total early deaths caused in each state, divided by the

population of the state. The right plots (c) show the total early deaths exported by each state, divided by the population of the state (that is, the difference between plots a and b). These impacts are based on summed contributions from each emitted species (see Fig. 3).

more early deaths outside of the state, than are caused within that state by emissions from elsewhere. A negative value indicates the opposite: that the state is a net importer of early deaths.

Three broad patterns are visible. First, the largest exporters are in the northern midwest, owing to low local populations, high emissions, and large downwind populations. Wyoming was the highest exporter on a per capita basis in 2005, with North Dakota and West Virginia following. While these states remained some of the largest per-capita exporters in 2018, their exported impacts fell by roughly 50% over this period (see the examples in Extended Data Table 2). Second, a cluster of states in the northeast are consistent net importers of impacts. New York was the highest net importer of early deaths in all three years, on both a per-capita and an absolute basis. For 2011, the approximately 2,800 deaths incurred in New York because of New York emissions represent 60% of the total deaths caused by New York emissions, and approximately 40% of the total air-quality-attributable deaths in the state. This implies that around 60% of deaths in New York are imported from other states. Finally, states on the west coast have a net exchange of around 0, owing to a combination of no upwind emissions (attributable to any state), relatively sparse population downwind, and large local populations. We present examples of state-level sectoral contributions in Extended Data Figs. 3–5.

Figures 3a, b present the US-wide early-death impacts for each sector and each chemical species, respectively. Impacts from all sectors decrease over the studied period, with the exception of commercial/residential and aviation (landing and take-off only). Impacts due to commercial/residential emissions increase by 31% between 2005 and 2011, but remain steady (within approximately 5%) from 2011 to 2018. Aviation landing and take-off impacts increase by approximately 60% between 2005 and 2018, but contribute around 0.3% to the summed 2018 impacts. Impacts from electric power generation reduce from 22% of total summed impacts in 2005 to 11% in 2018. We estimate that reductions in emissions from electric power generation have led to around 15,900 avoided early deaths in 2018 and, interpolating linearly, to approximately 137,000 avoided early deaths integrated over the 14 years analysed here. Because of these changes, electric power

generation changes from being the second most important emission sector to the fourth, while commercial/residential emissions go from fourth to first, responsible for 37% of the summed early deaths attributable to combustion emissions in 2018.

In terms of speciated impacts—that is, emissions species that contribute to the formation of, and exposure to, $\text{PM}_{2.5}$ and/or ozone—primary $\text{PM}_{2.5}$ emissions had the greatest impact in all three model years. They also stayed relatively consistent, with a 13% reduction in health impacts from 2005 to 2018. SO_2 —which was the third-greatest contributor to impacts in 2005, making up 19% of the summed impacts—was contributing less than 6% by 2018. This was due to an approximately 80% reduction in SO_2 emissions.

Ammonia-attributable impacts increased by around 21% between 2005 and 2018. This difference was driven by an increase in the sensitivity of $\text{PM}_{2.5}$ exposure with respect to a unit of ammonia emissions between 2005 and 2011. Owing to the decline in the importance of SO_2 , ammonia impacts went from being the fourth-greatest to the third-greatest contributor to total impacts over this period, increasingly close to the contribution of NO_x species. NO_x remained the second-greatest contributor to impacts from 2005 to 2018. Despite the roughly 50% reduction in total NO_x emissions between 2005 and 2018, impacts attributable to NO_x reduced by only around 35% between the two years. This is largely due to the increased sensitivity of $\text{PM}_{2.5}$ formation to NO_x emissions between 2005 and 2011, as noted previously²³.

On the basis of a linear combination of impacts by sector, we estimate US combustion emissions in 2005, 2011 and 2018 to have resulted in 111,200 (95% confidence interval 78,100–144,800), 93,700 (65,600–121,800) and 76,500 (53,300–99,600) early deaths, respectively. However, the total impact of all US anthropogenic emissions is different to the combined effect of each individual sector or species, owing to nonlinear interactions between the emitted chemicals (Fig. 3c). These interactions reduce the total impacts attributable to $\text{PM}_{2.5}$ by 30–34%. Impacts attributable to ozone instead increase by a factor of 2.4 to 2.8 (with the nonlinearity underlying this shown in Extended Data Fig. 6), raising the fraction of total early deaths attributable to ozone exposure from roughly 10% to around 30%.

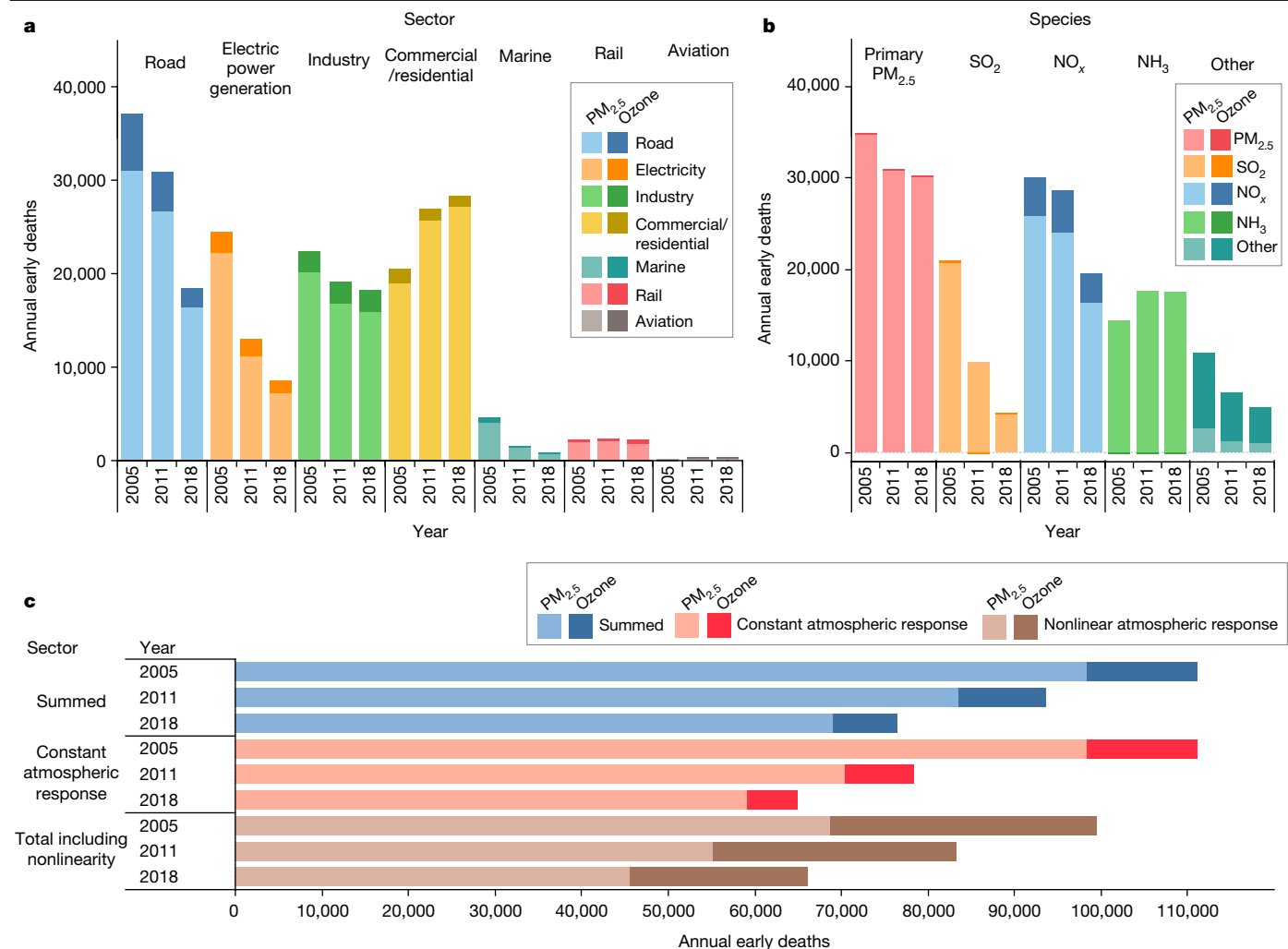


Fig. 3 | Total annual early deaths attributable to emission sector, emission species and in total. a, Total annual early deaths attributable to each emission sector. **b**, Total annual early deaths attributable to each emission species that leads to the formation of PM_{2.5} and/or ozone. **c**, Total annual early deaths. Data are shown for 2005, 2011 and 2018, and for PM_{2.5} and ozone. In **c**, three totals are

presented: the sum of all sectors/species ('Summed'), which does not account for nonlinear interactions between species; the sum of all sectors/species with varying emissions, but constant (2005) atmosphere ('Constant atmospheric response'); and the total impacts after accounting for nonlinear interactions between species. Tabulated results are presented in Extended Data Table 3.

Taking these nonlinearity effects into account results in total US combustion emissions impacts of 96,600 (95% confidence interval 74,200–125,000), 83,300 (62,400–104,200) and 66,100 (49,300–82,900) early deaths for 2005, 2011 and 2018, respectively. This effect highlights the difference in expected changes in population exposure that result from marginal changes by comparison with larger-scale emissions increases or reductions. An explanation of this effect and its quantification is given in the Methods. The atmospheric nonlinearity is also reflected in our computed sensitivity differences between 2005 and 2011. Thus, a 1% reduction in 2011 emissions would lead to roughly 940 avoided early deaths. Had the atmospheric response to a unit of emissions remained constant between 2005 and 2011 (in terms of meteorology and background concentrations), the same emissions reduction would have led to around 780 avoided early deaths. The changing atmospheric composition thus increases the early deaths attributable to a unit of emission. These three effects are displayed in Fig. 3c.

Overall, we have found that more than 40% of the combustion-emissions-related early deaths cross state lines. This highlights the need for a cooperative approach between states for reaching air-quality targets or targeting problematic areas, as underlined by the introduction of EPA's CSAPR⁵. We find that the electric power generation sector is of

declining importance to air quality, by comparison with the increasing importance of commercial/residential emissions. A 10% decrease in emissions from the commercial/residential sector would have 3.3 times greater benefit than a further 10% decrease in emissions from electric power generation. This is reflected in the declining relative importance of SO₂, and the increasing relative importance of primary PM_{2.5} and ammonia. A 10% decrease in primary PM_{2.5}, NO_x and ammonia emissions would now have 7, 4.5 and 4 times the benefit, respectively, compared with a further 10% decrease in SO₂ emissions. These changing relative sectoral and speciated influences provide room to advance air-quality mitigation efforts in the US.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1983-8>.

1. World Health Organization *Health Risks of Particulate Matter from Long-range Transboundary Air Pollution* (WHO Regional Office for Europe, 2006).

2. United States Environmental Protection Agency *The Benefits and Costs of the Clean Air Act from 1990 to 2020* Final Report of US Environmental Protection Agency (Office of Air and Radiation, 2011).
3. Landrigan, P. J. et al. The Lancet Commission on pollution and health. *Lancet* **391**, 462–512 (2018).
4. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525**, 367–371 (2015).
5. United States Environmental Protection Agency *Regulatory Impact Analysis for the Federal Implementation Plans to Reduce Interstate Transport of Fine Particulate Matter and Ozone in 27 States; Correction of SIP Approvals for 22 States* Report EPA-HQ-OAR-2009-0491 (Office of Air and Radiation, 2011).
6. Goodkind, A. L., Tessum, C. W., Coggins, J. S., Hill, J. D. & Marshall, J. D. Fine-scale damage estimates of particulate matter air pollution reveal opportunities for location-specific mitigation of emissions. *Proc. Natl Acad. Sci. USA* **116**, 8775–8780 (2019).
7. Krewski, D. et al. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Res. Rep. Health. Eff. Inst.* **140**, 5–114 (2009).
8. Lim, S. S. et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2224–2260 (2012).
9. Hoek, G. et al. Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environ. Health* **12**, 43 (2013).
10. Burnett, R. T. et al. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* **122**, 397–403 (2014).
11. Forouzanfar, M. H. et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 2287–2323 (2015).
12. Burnett, R. et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl Acad. Sci. USA* **115**, 9592–9597 (2018).
13. Caiazzo, F., Ashok, A., Waitz, I. A., Yim, S. H. L. & Barrett, S. R. H. Air pollution and early deaths in the United States. Part I: quantifying the impact of major sectors in 2005. *Atmos. Environ.* **79**, 198–208 (2013).
14. Fann, N. et al. Estimating the national public health burden associated with exposure to ambient PM_{2.5} and ozone. *Risk Anal. Off. Publ. Soc. Risk Anal.* **32**, 81–95 (2012).
15. Penn, S. L. et al. Estimating state-specific contributions to PM_{2.5}- and O₃-related health burden from residential combustion and electricity generating unit emissions in the United States. *Environ. Health Perspect.* **125**, 324–332 (2017).
16. Turner, M. D. et al. Premature deaths attributed to source-specific BC emissions in six urban US regions. *Environ. Res. Lett.* **10**, 114014 (2015).
17. Tong, D. Q. & Mauzerall, D. L. Summertime state-level source-receptor relationships between nitrogen oxides emissions and surface ozone concentrations over the continental United States. *Environ. Sci. Technol.* **42**, 7976–7984 (2008).
18. Fann, N., Fulcher, C. M. & Baker, K. The recent and future health burden of air pollution apportioned across U.S. sectors. *Environ. Sci. Technol.* **47**, 3580–3589 (2013).
19. Dedoussi, I. C. & Barrett, S. R. H. Air pollution and early deaths in the United States. Part II: attribution of PM_{2.5} exposure to emissions species, time, location and sector. *Atmos. Environ.* **99**, 610–617 (2014).
20. Pappin, A. J. & Hakami, A. Source attribution of health benefits from air pollution abatement in Canada and the United States: an adjunct sensitivity analysis. *Environ. Health Perspect.* **121**, 572–579 (2013).
21. Fann, N., Kim, S.-Y., Olives, C. & Sheppard, L. Estimated changes in life expectancy and adult mortality resulting from declining PM_{2.5} exposures in the contiguous United States: 1980–2010. *Environ. Health Perspect.* **125**, 097003 (2017).
22. Henze, D. K., Hakami, A. & Seinfeld, J. H. Development of the adjunct of GEOS-Chem. *Atmos. Chem. Phys.* **7**, 2413–2433 (2007).
23. Holt, J., Selin, N. E. & Solomon, S. Changes in inorganic fine particulate matter sensitivities to precursors due to large-scale US emissions reductions. *Environ. Sci. Technol.* **49**, 4834–4841 (2015).
24. Bureau of Economic Analysis *BEA Regions* <http://www.bea.gov/regional/docs/regions.cfm> (2004).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

We present here the data and models used in calculating the cross-state early-deaths caused by combustion emissions. We first estimate the speciated emissions for each combustion sector. We then use the adjoint of a chemistry-transport model to estimate the impact of changes in emissions on population exposure. Finally, we relate increases in population exposure to public health impacts (early deaths) using epidemiologically derived concentration-response functions. These steps, intercomparison of our results against existing literature, and the limitations of our approach are outlined below.

Combustion emissions

Emissions are attributed to each of the six nonaviation sectors in the US using SMOKE and the EPA National Emissions Inventory (NEI) for the year of 2005 (previously used in ref. ¹³), 2011 (NEI2011v6 version 1) and the 2011-based forecast for 2018 (refs. ^{25–27}). These are generated on a 12 km × 12 km or 36 km × 36 km grid, and regridded to the 0.5° × 0.666° (latitude × longitude) grid of the nested GEOS-Chem adjoint model. The full list of individual sources (and corresponding EPA source classification code (SCC) identifiers) that comprise each sector are provided in the data repository noted in the Methods section ‘Data and code availability’. For road transportation for 2011 and 2018, we use the EPA Motor Vehicle Emissions Simulator (MOVES)-processed emissions²⁸. For aviation emissions we use the Aviation Environmental Design Tool (AEDT) inventories for 2006, 2010, 2012 and 2015 (ref. ²⁹). When referring to each of 2005, 2011 and 2018 aviation impacts, we imply impacts from 2006, the average of 2010 and 2012, and 2015 emissions respectively, owing to the absence of more recent datasets. Only aviation emissions that occur within 1 km of the surface (landing and take-off emissions) are taken into account. These have been shown to capture roughly one-third of total aviation-emissions-attributable early deaths in the US^{30,31}. We account for the underrepresentation of EPA’s point-source oil and gas sector (pt_oilgas) in NEI2011v6 version 1, by distributing the underrepresented NO_x (the difference in pt_oilgas NO_x between version 3 and version 1) to the industry sector NO_x emissions on a state level, assuming the existing spatial distribution²⁷. When calculating state source–receptor matrices for the marine sector, we only consider marine emissions within state boundaries and within, on average, around 25 km off the coast over the sea (where applicable). Besides the marine sector, which does not necessarily fall within state boundaries, we do not account for the impacts of emissions that occur outside of this domain and might contribute to US early deaths. Further details on emissions modelling are provided in the data repository.

Air-quality modelling

We use the adjoint of the GEOS-Chem chemistry-transport model²² to calculate the sensitivities of the aggregate population exposure in each of the 48 contiguous US states with respect to the various emission species in the North American domain. The resolution of the horizontal grid is 0.5° × 0.666° (roughly 55 km × 55 km) (latitude × longitude), with 47 vertical layers up to 80 km. This horizontal resolution is adequate for capturing state-wide impacts^{32–34}. Boundary conditions for the nested domain are obtained from the global GEOS-Chem model run at 4° × 5° resolution, driven by corresponding global meteorological data. Each of the 48 sensitivities quantifies the effect that any emission species in any location in the contiguous US and at any time will have on the population exposure to PM_{2.5} or ozone in each corresponding state. We define PM_{2.5} as the mass sum of nitrates, sulfates, ammonium, black carbon and organic carbon, capturing both primary and secondary PM_{2.5} concentrations. Secondary organic aerosols are not captured. We perform an annual simulation for each of PM_{2.5} and ozone state-level exposure, in each contiguous US state, for 2006 and 2011, resulting in 192 annual adjoint simulations in total (48 × 2 × 2). We use GEOS assimilated meteorological data from the Global Modelling and Assimilation

Office (GMAO) at the NASA Goddard Space Flight Center. The year 2006 was climatologically warm in the US, with the annual average temperature being 0.55 °C higher than the 1995–2015 mean, whereas 2011 was climatologically average with an average temperature 0.04 °C lower than the 1995–2015 mean³⁵. For 2018 we use the 2011 atmospheric response. Given the change between 2005 and 2011 (comparing the ‘Summed’ and ‘Constant atmospheric response’ in Fig. 3), we expect that this approximation will result in a maximum error of around 15% (as there were larger emissions changes between 2005 and 2011 than between 2011 and 2018). Total impacts across all sectors are calculated using additional ‘forward’ runs, described at the end of this section.

The GEOS-Chem baseline emissions are from EPA’s NEI for 2005 and 2011 accordingly^{26,27}. Previous studies have found that the NEI 2011 road transportation NO_x emissions are overestimated by around 50% in the southeast and nationally^{36,37}. The effects of this are not included here as they are, as of the time of writing, not incorporated in EPA’s NEI. An overestimation of 50% in the road transportation NO_x emissions in 2011 implies that results presented here overestimate road transportation early deaths by around 7,500 (95% confidence interval 5,200–9,700) early deaths per year. Other emissions sources, both natural and anthropogenic, are simulated using the standard GEOS-Chem nested North American domain datasets. The Electronic Data Gathering, Analysis and Retrieval (EDGAR) global anthropogenic emissions inventory drives the global model (from which the boundary conditions for the nested simulations are generated)³⁸. This is replaced by regional emissions inventories where available (for example, NEI). Biogenic emissions are from the Model of Emissions of Gases and Aerosols from Nature (MEGAN) inventory³⁹, and lighting NO_x emissions are calculated on the basis of ref. ⁴⁰.

We estimate the impacts of each sector by performing an inner (Hadamard) product of the sensitivities with the gridded emissions for each of the seven sectors, and calculate the corresponding population exposure impacts. This linear approach was used and validated in refs. ^{19,20,41–43} against the forward model difference method.

When calculating the total impacts from all sectors combined, we use a different approach to take into account nonlinear interactions between the sectors. Total impacts are calculated by comparing the surface concentrations in forward GEOS-Chem simulations with and without all US anthropogenic emissions. These forward model simulations allow us to quantify nonlinearity in the response of US air quality. Sets of seven forward simulations are conducted for both 2005 and 2011 to quantify this nonlinearity. Extended Data Fig. 6 shows how the simulated, population-weighted concentrations of ozone and PM_{2.5} respond to large changes in emissions (‘Average sensitivity’). Compared with the sensitivities used for single-sector and speciated impact calculations (‘Marginal sensitivity’), the full, nonlinear PM_{2.5} response to removal of all emissions is found to be 30–34% smaller, while the ozone response is found to be 2.4–2.8 times greater, implying greater nonlinearity effects for ozone by comparison with PM_{2.5}. This is because ozone sensitivities are larger when ozone concentrations are low, owing to the greater ozone-production efficiencies in a clean background atmosphere⁴⁴. For PM_{2.5}, the response nonlinearity is driven by competition between SO₄ (from emitted SO₂) and NO₃ (from emitted NO_x) for ammonia^{23,45}.

Total impacts for 2018 are estimated by scaling the 2011 response. The scaling factor is calculated as the total growth in US population, multiplied by the ratio of the linearized response to 2018 and 2011 emissions.

Health impacts

We quantify air-quality impacts in terms of early deaths (premature mortalities). The toxicity of different PM_{2.5} species is assumed to be equal, consistent with EPA practice. As with any study of air pollution impacts, our results are sensitive to the specific choice of concentration–response function (CRF). To calculate the effects of PM_{2.5} exposure, we apply the American Cancer Society (ACS) cohort study

log-linear response estimate of 6% (range 4–8%) increased risk of all-cause mortality per $10 \mu\text{g m}^{-3}$ increase in annually averaged $\text{PM}_{2.5}$ exposure, derived for 1999–2000 exposures using the random-effects Cox model, and adjusted for 44 individual-level and 7 ecological covariates⁷. This estimate is linearized and applied here for adults over the age of 30 years old. This CRF has been applied in a number of estimates of US pollution impacts^{46–48}; it is consistent with the results of a global meta-analysis of epidemiological literature, which also found a 6% (range 4–8%) increase in risk per $10 \mu\text{g m}^{-3}$ (ref. ⁹).

Using a different risk estimate would result in a change in the total estimated impact. An expert elicitation performed by the EPA indicated a 1% (range 0.4–1.8%) increase in all-cause mortalities per $1 \mu\text{g m}^{-3}$ of exposure². This would imply a roughly 70% increase in calculated early deaths, although all relative comparisons would remain the same. Another alternative based on the US medicare cohort would imply a roughly 18% increase in the calculated early deaths for $\text{PM}_{2.5}$, when applied to the same 30-plus population (again with all relative comparisons staying the same, but with the caveat that this was derived in a 65-plus cohort)⁴⁹. Extended Data Table 4 shows how the estimate of total impacts, accounting for nonlinearity of the atmospheric response, is affected by the estimated relative risk, including the previously cited studies^{2,7,12}, refs. ^{49–51} and the results of a meta-analysis of epidemiological literature⁹. Although we cannot directly apply a nonlinear CRF, using the mean 2011 US concentration of $\text{PM}_{2.5}$ in the global exposure mortality model (GEMM)¹², we estimate a 35% increase in calculated early deaths.

For ozone, we apply the respiratory disease mortality CRF of ref. ⁵²; this is based on US exposure data from the same ACS study as above⁷. Impacts are calculated using the 8-hour maximum daily average ozone over the entire year, and applied to the same population. However, as with $\text{PM}_{2.5}$, there is disagreement regarding the correct exposure response curve to use. Extended Data Table 4 also includes estimates of ozone impacts, accounting for nonlinearity of the atmospheric response, using different ozone exposure response curves from the literature^{50,52,53}. Using the all-cause mortality CRF of ref. ⁵² would result in a 110% increase in total mortality due to ozone exposure. Applying the all-cause mortality CRF of ref. ⁵⁰ to quantify ozone health impacts would instead result in a roughly 17% increase in the reported early deaths due to ozone exposure. We note that the CRF of ref. ⁵⁰ is based on mean summertime ozone exposure, whereas we measure annual-average exposure to 8-hour maximum ozone. However, ref. ⁵² showed that the response of respiratory mortality to chronic ozone exposure is similar when using either annual average (12% increase per 10 ppbv) or warm season (10% per 10 ppbv) exposure.

Population data are obtained from the global rural urban mapping project (GRUMP)⁵⁴ and LandScan⁵⁵ databases. For 2018, we scale the 2011 population to match the 2017 US Census totals⁵⁶. State population fractions over the age of 30 years old are obtained from the US Census Bureau for 2011 (ref. ⁵⁷). The US baseline all-cause and respiratory disease incidence rates are obtained from the WHO for 2012 (ref. ⁵⁸). For both $\text{PM}_{2.5}$ and ozone, the early-deaths confidence intervals reflect the reported uncertainty range for the CRF. Uncertainty in the summed $\text{PM}_{2.5}$ and ozone impacts is calculated by performing a Monte Carlo simulation with 10^6 independent draws of each CRF, applying a triangular distribution to both.

Intercomparison with other studies

Pollution exchange on an intercontinental scale has previously been estimated for ozone^{59–61}, $\text{PM}_{2.5}$ (refs. ^{62–65}), and both⁶⁶, highlighting the influence of emissions from cross-continental sources. Regional studies have focused on individual species or species and pollutants—for example, the NO_x to ozone effect between EU countries⁶⁷ and between US states¹⁷, sources of black-carbon impacts in parts of the US¹⁶, and fine-scale monetized US $\text{PM}_{2.5}$ impacts of different sectors⁶, in addition

to other studies not using detailed chemistry-transport model (CTM) approaches.

The main contribution of our work is the breakdown of both air-pollution causes and impacts in the US, and there are no studies to which direct comparisons at the level of disaggregation in our work can be made. However, the aggregate results of this study compare well with those in the existing literature. Ref. ⁶⁸ reports a roughly 25% decrease in $\text{PM}_{2.5}$ -attributable early deaths in the US between 2005 and 2014, which is similar to the roughly 22% found here (interpolating for these two years). Our estimated total early deaths fall within the uncertainty ranges of recent studies, for example, the 79,300 (95% confidence interval 39,700–113,000) non-agriculture-related 2015 US early deaths reported in ref. ⁶⁹; the 88,400 (66,800–115,000) 2015 US $\text{PM}_{2.5}$ -attributable early deaths reported in ref. ⁷⁰; and the central estimate of 107,000 total 2011 US $\text{PM}_{2.5}$ -attributable deaths (of which around 85,600 correspond to non-agriculture- and non-fire-related deaths) reported in ref. ⁶. As in these studies, our 2011 estimates are higher than the 2010 estimates of ref. ⁴ (around 37,400 US early deaths for non-natural and non-agriculture-related deaths). In addition, refs. ^{4,69} report different sectoral attributions, probably owing to the different emissions inventory used (EDGAR versus NEI). Our sectoral and speciated relative attribution is similar (for 2005) to that of ref. ¹⁹ (with the absolute values being different because of the different health-impacts function applied).

We also compare our estimated changes in population exposure to data obtained from monitor sites. We find that, between 2005 and 2011, the simulated population exposure to $\text{PM}_{2.5}$ and ozone (taking into account nonlinearities) fell by roughly 20% and 8.6% respectively. For the same two years, EPA's annual trends from nationwide monitor sites show a decrease of 24% and 8% for $\text{PM}_{2.5}$ and ozone concentrations respectively⁷¹.

Limitations

In terms of air-quality modelling, even though the $0.5^\circ \times 0.666^\circ$ (roughly $55 \text{ km} \times 55 \text{ km}$) (latitude \times longitude) resolution is sufficient for capturing state-level regional impacts, it may underestimate primary $\text{PM}_{2.5}$ impacts and misrepresent ozone impacts in densely populated urban areas. This is in part due to the instantaneous dilution of the emissions, and, for ozone, to the highly nonlinear relationship between ozone formation and background VOC and NO_x concentrations. The EPA NEIs that are used here, and in policy assessments, are also only an approximation, with some known issues that we do not explicitly account for^{36,37}. This could affect both the baseline calculation of the sensitivity and the absolute impacts attribution. In addition, the emissions presented for 2018 are forecasted from the NEI2011 inventory. Such forecasts are inherently uncertain^{72–74}. Finally, previous studies have shown a tendency for GEOS-Chem simulations to overestimate nitrates^{75,76}. This may result in artificially increased $\text{PM}_{2.5}$ formation in response to combustion emissions.

In estimating health impacts, the choice of CRF is critical for early-death calculations. Here we apply the all-cause CRF for $\text{PM}_{2.5}$ from the ACS cohort study⁷ because of the large and nationally representative cohort it is based on, and because of its wide application in $\text{PM}_{2.5}$ -attributable health-impact estimates in the literature. This CRF was derived for pre-2000 concentrations, and we thus assume no heterogeneity in effect estimates over time (as concentrations change). An analysis of the level of disagreement between different CRFs, and the effect on our estimated impacts, is presented in the 'Health impacts' section above.

We assume equal toxicity between different $\text{PM}_{2.5}$ species, consistent with EPA's practice. However, epidemiological work on differential toxicity has provided estimates for mortality predictors based on exposure to individual $\text{PM}_{2.5}$ constituents⁷⁷. Sulfates and black carbon have specifically been highlighted because of their suspected higher toxicity amongst $\text{PM}_{2.5}$ constituents^{9,78}.

Here we choose to quantify all-cause and respiratory-disease mortality for long-term exposure to PM_{2.5} and ozone respectively, but note that human exposure to PM_{2.5} and ozone has been correlated with a variety of specific health endpoints, such as neurological diseases⁷⁹, various forms of cancer⁸⁰, low birth weight⁸¹, and others. Short-term exposure to PM_{2.5} and ozone has also been found to correlate causally with an increased likelihood of early death^{82,83}, and is not included here. Nonfatal (morbidity) effects attributable to PM_{2.5} and ozone exposure—including acute respiratory symptoms, exacerbated asthma, days of work and school lost, upper and lower respiratory symptoms, nonfatal heart attacks, acute bronchitis, and hospital and emergency-department visits—are also not captured. In addition, given the aggregate nature of the adjoint objective function, we present results for the aggregate state-level population. Air-pollution-related health impacts, however, have been known to disproportionately affect different races, ages and socioeconomic backgrounds^{84,85}. These are not broken down here.

We also note that this work quantifies the pollution exchange between the contiguous US states, and does not take into account sources outside of this domain (for example, Mexico, Canada and intercontinentally^{65,86}). In addition, while changes in emissions are probably the largest driver of changes in the cross-state, sectoral and speciated patterns between the years, effects of meteorological changes can also contribute, and are not specifically decoupled here. Finally, for simultaneous, large changes in multiple pollutant emissions, there may be nonlinear interactions. These interactions could change the total impact relative to that calculated for individual sectors here, where independent changes are assumed. For this reason, and as discussed above, we calculate and present total impacts (aggregated across all sectors) using forward simulations in which all emissions are reduced simultaneously.

Data availability

The cross-state source–receptor matrices generated and analysed here, together with sector definitions, are available in the 4TU.ResearchData repository at <https://doi.org/10.4121/uuid:edfc5304-39ed-4556-a95a-f8b3313f7cfc>.

Code availability

The atmospheric modelling code used is publicly available; instructions for download are given at http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint.

25. United States Environmental Protection Agency *SMOKE v3.5.1 User's Manual* (Institute for the Environment, University of North Carolina at Chapel Hill, 2013).
26. United States Environmental Protection Agency *2005 National Emissions Inventory (NEI) data*. Environmental Protection Agency Air Emissions Inventories. <https://www.epa.gov/air-emissions-inventories/2008-national-emissions-inventory-nei-data> (2008).
27. United States Environmental Protection Agency *2011 National Emissions Inventory*. Environmental Protection Agency Air Emissions Inventories. <https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data> (2014).
28. United States Environmental Protection Agency *MOVES and Other Mobile Source Emissions Models* <https://www.epa.gov/moves> (Environmental Protection Agency, 2017).
29. Wilkerson, J. T. et al. Analysis of emission data from global commercial aviation: 2004 and 2006. *Atmos. Chem. Phys.* **10**, 6391–6408 (2010).
30. Yim, S. H. L. et al. Global, regional and local health impacts of civil aviation emissions. *Environ. Res. Lett.* **10**, 034001 (2015).
31. Eastham, S. D. & Barrett, S. R. H. Aviation-attributable ozone as a driver for changes in mortality related to air quality and skin cancer. *Atmos. Environ.* **144**, 17–23 (2016).
32. Thompson, T. M., Saari, R. K. & Selin, N. E. Air quality resolution for health impact assessment: influence of regional characteristics. *Atmos. Chem. Phys.* **14**, 969–978 (2014).
33. Li, Y., Henze, D. K., Jack, D. & Kinney, P. L. The influence of air quality model resolution on health impact assessment for fine particulate matter and its components. *Air Qual. Atmos. Health* **9**, 51–68 (2016).
34. Arunachalam, S., Wang, B., Davis, N., Baek, B. H. & Levy, J. I. Effect of chemistry-transport model scale and resolution on population exposure to PM_{2.5} from aircraft emissions during landing and takeoff. *Atmos. Environ.* **45**, 3294–3300 (2011).

35. National Oceanic and Atmospheric Administration *Climate at a Glance* <https://www.ncdc.noaa.gov/cag/> (National Centers for Environmental Information, 2017).
36. Travis, K. R. et al. Why do models overestimate surface ozone in the Southeast United States? *Atmos. Chem. Phys.* **16**, 13561–13577 (2016).
37. Anderson, D. C. et al. Measured and modeled CO and NO_x in DISCOVER-AQ: an evaluation of emissions and chemistry over the eastern US. *Atmos. Environ.* **96**, 78–87 (2014).
38. Olivier, J. G. J. et al. *Applications of EDGAR. Including a Description of EDGAR 3.2 Reference Database with Trend Data for 1970–1995*. RIVM Report 773301 001/NRP Report 410200 051 (2002).
39. Guenther, A. B. et al. The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geosci. Model Dev.* **5**, 1471–1492 (2012).
40. Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C. & Koshak, W. J. Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data. *J. Geophys. Res. Atmos.* **117**, D20307 (2012).
41. Barrett, S. R. H. et al. Impact of the Volkswagen emissions control defeat device on US public health. *Environ. Res. Lett.* **10**, 114005 (2015).
42. Lee, C. J. et al. Response of global particulate-matter-related mortality to changes in local precursor emissions. *Environ. Sci. Technol.* **49**, 4335–4344 (2015).
43. Ashok, A. & Barrett, S. R. H. Adjoint-based computation of U.S. nationwide ozone exposure isopleths. *Atmos. Environ.* **133**, 68–80 (2016).
44. Wu, S., Duncan, B. N., Jacob, D. J., Fiore, A. M. & Wild, O. Chemical nonlinearities in relating intercontinental ozone pollution to anthropogenic emissions. *Geophys. Res. Lett.* **36**, L05806 (2009).
45. Pinder, R. W., Adams, P. J. & Pandis, S. N. Ammonia emission controls as a cost-effective strategy for reducing atmospheric particulate matter in the eastern United States. *Environ. Sci. Technol.* **41**, 380–386 (2007).
46. Anenberg, S. C., Horowitz, L. W., Tong, D., Daniel, Q. & West, J. J. An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environ. Health Perspect.* **118**, 1189–1195 (2010).
47. Evans, J. et al. Estimates of global mortality attributable to particulate air pollution using satellite imagery. *Environ. Res.* **120**, 33–42 (2013).
48. Wolfe, P. et al. Monetized health benefits attributable to mobile source emission reductions across the United States in 2025. *Sci. Total Environ.* **650**, 2490–2498 (2019).
49. Lepeule, J., Laden, F., Dockery, D. & Schwartz, J. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard six cities study from 1974 to 2009. *Environ. Health Perspect.* **120**, 965–970 (2012).
50. Di, Q. et al. Air pollution and mortality in the Medicare population. *N. Engl. J. Med.* **376**, 2513–2522 (2017).
51. Pope, C. A. et al. Mortality risk and PM_{2.5} air pollution in the USA: an analysis of a national prospective cohort. *Air Qual. Atmos. Health* **11**, 245–252 (2018).
52. Turner, M. C. et al. Long-term ozone exposure and mortality in a large prospective study. *Am. J. Respir. Crit. Care Med.* **193**, 1134–1142 (2016).
53. Jerrett, M. et al. Long-term ozone exposure and mortality. *N. Engl. J. Med.* **360**, 1085–1095 (2009).
54. Balk, D. L. et al. Determining global population distribution: methods, applications and data. *Adv. Parasitol.* **62**, 119–156 (2006).
55. Bright, E. A., Coleman, P. R., Rose, A. N. & Urban, M. L. *LandScan 2013 High Resolution Global Population Data Set v.13* (United States Department of Energy Office of Scientific and Technical Information, 2014).
56. United States Census Bureau *Annual estimates of the resident population: April 1, 2010 to July 1, 2017* <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk> (USCB, 2018).
57. United States Census Bureau. 2007–2011 American community survey 5-year estimates: age and sex. *American Fact Finder* <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml> (2011).
58. World Health Organization *Disease and Injury Country Estimates, 2000–2012* http://www.who.int/healthinfo/global_burden_disease/estimates_country_2000_2012/en/ (WHO, 2014).
59. Fiore, A. M. et al. Multimodel estimates of intercontinental source-receptor relationships for ozone pollution. *J. Geophys. Res. Atmos.* **114**, D04301 (2009).
60. West, J. J., Naik, V., Horowitz, L. W. & Fiore, A. M. Effect of regional precursor emission controls on long-range ozone transport—part 2: steady-state changes in ozone air quality and impacts on human mortality. *Atmos. Chem. Phys.* **9**, 6095–6107 (2009).
61. Anenberg, S. C. et al. Intercontinental impacts of ozone pollution on human mortality. *Environ. Sci. Technol.* **43**, 6482–6487 (2009).
62. Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Dingenen, R. V. & Dentener, F. Contribution and uncertainty of sectorial and regional emissions to regional and global PM_{2.5} health impacts. *Atmos. Chem. Phys.* **19**, 5165–5186 (2019).
63. Liu, J., Mauzerall, D. L. & Horowitz, L. W. Evaluating inter-continental transport of fine aerosols: (2) global health impact. *Atmos. Environ.* **43**, 4339–4347 (2009).
64. Anenberg, S. C. et al. Impacts of intercontinental transport of anthropogenic fine particulate matter on human mortality. *Air Qual. Atmos. Health* **7**, 369–379 (2014).
65. Zhang, Q. et al. Transboundary health impacts of transported global air pollution and international trade. *Nature* **543**, 705–709 (2017).
66. Liang, C.-K. et al. HTAP2 multi-model estimates of premature human mortality due to intercontinental transport of air pollution and emission sectors. *Atmos. Chem. Phys.* **18**, 10497–10520 (2018).
67. Lupascu, A. & Butler, T. Source attribution of European surface O₃ using a tagged O₃ mechanism. *Atmos. Chem. Phys.* **19**, 14535–14558 (2019).
68. Fann, N., Coffman, E., Timin, B. & Kelly, J. T. The estimated change in the level and distribution of PM_{2.5}-attributable health impacts in the United States: 2005–2014. *Environ. Res.* **167**, 506–514 (2018).
69. Lelieveld, J. Clean air in the Anthropocene. *Faraday Discuss.* **200**, 693–703 (2017).

70. Cohen, A. J. et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389**, 1907–1918 (2017).
71. United States Environmental Protection Agency *National Air Quality: Status and Trends of Key Air Pollutants* <https://www.epa.gov/air-trends> (US EPA, 2018).
72. Tong, D. Q. et al. Long-term NO_x trends over large cities in the United States during the great recession: comparison of satellite retrievals, ground observations, and emission inventories. *Atmos. Environ.* **107**, 70–84 (2015).
73. Tong, D. Q., Lee, P. & Saylor, R. D. New directions: the need to develop process-based emission forecasting models. *Atmos. Environ.* **47**, 560–561 (2012).
74. Jiang, Z. et al. Unexpected slowdown of US pollutant emission reduction in the past decade. *Proc. Natl Acad. Sci. USA* **115**, 5099–5104 (2018).
75. Heald, C. L. et al. Atmospheric ammonia and particulate inorganic nitrogen over the United States. *Atmos. Chem. Phys.* **12**, 10295–10312 (2012).
76. Zhu, L. et al. Sources and impacts of atmospheric NH₃: current understanding and frontiers for modeling, measurements, and remote sensing in North America. *Curr. Pollut. Rep.* **1**, 95–116 (2015).
77. Crouse, D. L. et al. A new method to jointly estimate the mortality risk of long-term exposure to fine particulate matter and its components. *Sci. Rep.* **6**, 18916 (2016).
78. Beelen, R. et al. Natural-cause mortality and long-term exposure to particle components: an analysis of 19 European cohorts within the multi-center ESCAPE project. *Environ. Health Perspect.* **123**, 525–533 (2015).
79. Kioumourtoglou, M.-A. et al. Long-term PM_{2.5} exposure and neurological hospital admissions in the northeastern United States. *Environ. Health Perspect.* **124**, 23–29 (2016).
80. Turner, M. C. et al. Ambient air pollution and cancer mortality in the cancer prevention study II. *Environ. Health Perspect.* **125**, 087013 (2017).
81. Rosa, M. J. et al. Prenatal exposure to PM_{2.5} and birth weight: a pooled analysis from three North American longitudinal pregnancy cohort studies. *Environ. Int.* **107**, 173–180 (2017).
82. Joel, S., Kelvin, F. & Antonella, Z. A national multicity analysis of the causal effect of local pollution, NO₂, and PM_{2.5} on mortality. *Environ. Health Perspect.* **126**, 087004 (2018).
83. Di, Q. et al. association of short-term exposure to air pollution with mortality in older adults. *J. Am. Med. Assoc.* **318**, 2446–2456 (2017).
84. Wang, Y. et al. Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. *Epidemiology* **28**, 771–779 (2017).
85. Kioumourtoglou, M.-A., Schwartz, J., James, P., Dominici, F. & Zanobetti, A. PM_{2.5} and mortality in 207 US cities: modification by temperature and city characteristics. *Epidemiology* **27**, 221–227 (2016).
86. Lin, J. et al. China's international trade and air pollution in the United States. *Proc. Natl Acad. Sci. USA* **111**, 1736–1741 (2014).

Acknowledgements We thank the EPA and K. Travis (Harvard) for providing assistance with the NEI datasets. This publication was made possible by US EPA grant RD-835872-01. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. I.C.D. was additionally funded through the Massachusetts Institute of Technology (MIT) Martin Family Fellowship for Sustainability and the MIT George and Marie Vergottis Fellowship. We also acknowledge support by the VoLo Foundation.

Author contributions I.C.D. and S.R.H.B. planned the research. I.C.D. performed the emissions modelling and the air quality modelling PM_{2.5} simulations. S.D.E. and E.M. performed the air quality modelling ozone simulations. I.C.D. and S.D.E. performed results analysis. I.C.D. drafted the manuscript with the help of S.D.E. and S.R.H.B. All authors provided feedback on the manuscript.

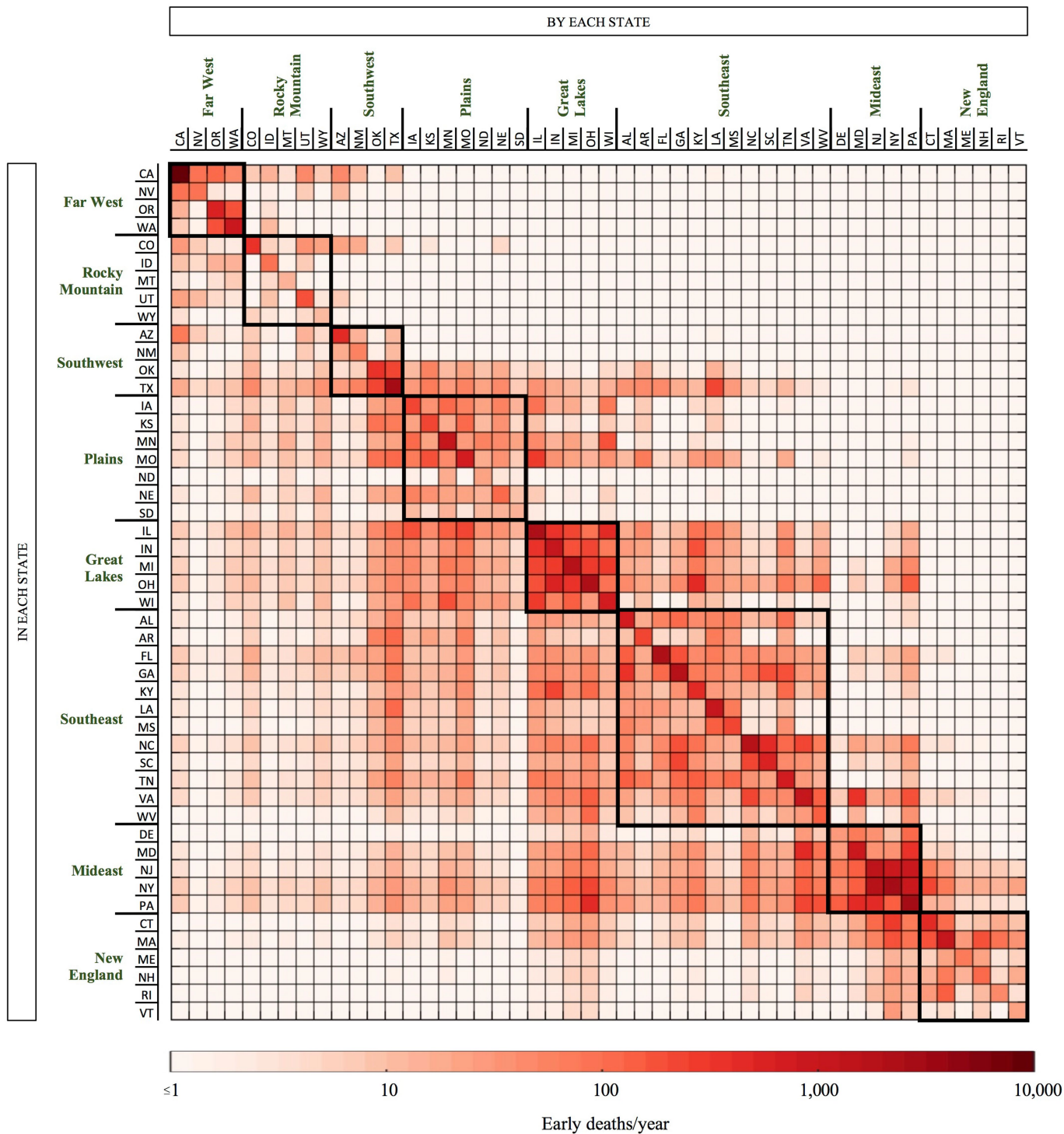
Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.R.H.B.

Peer review information Nature thanks Marianthi-Anna Kioumourtoglou, Enrico Pisoni, Andrea Pozzer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

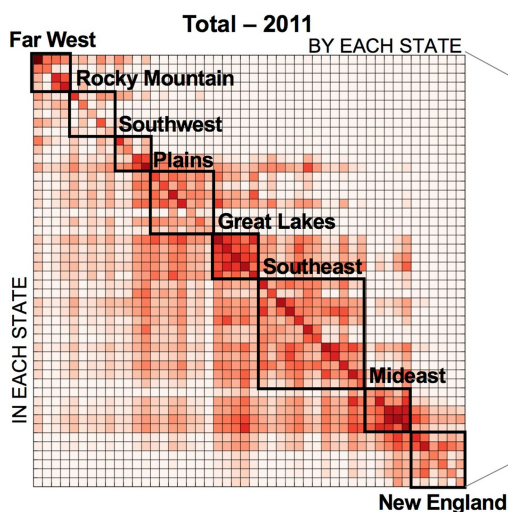
Reprints and permissions information is available at <http://www.nature.com/reprints>.



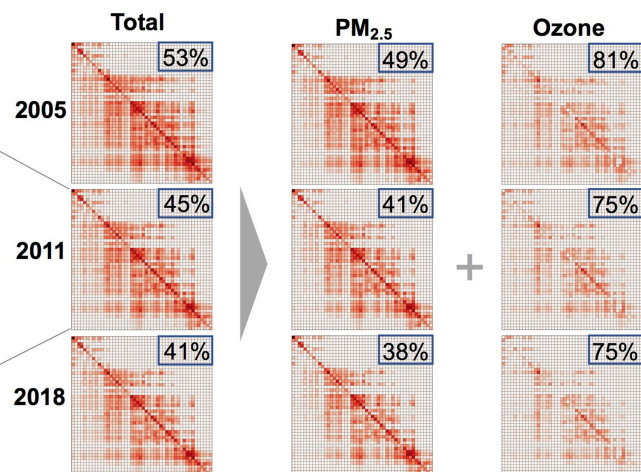
Extended Data Fig. 1 | Source-receptor matrix showing total impacts in 2011 for the contiguous US. 'By each state' indicates sources; 'in each state' indicates receptors. The matrix is annotated with state abbreviations and their regional grouping.

(a) Source-receptor of total early deaths

(i) 2011 – total

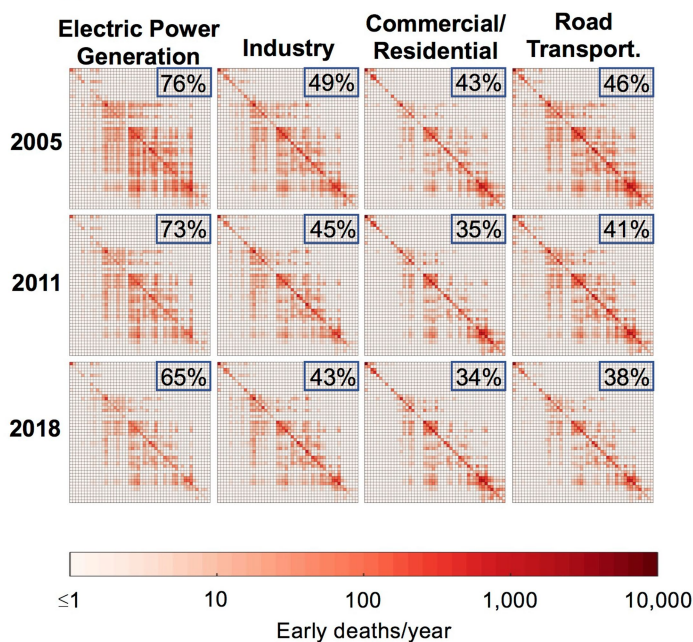


(ii) 2005-2018 – total, PM_{2.5}, and ozone

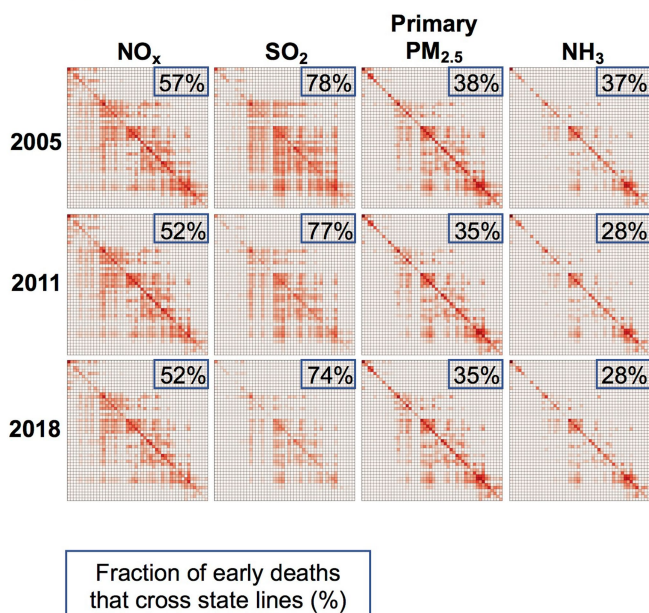


(b) Source-receptor early death attribution

(i) Emission sector



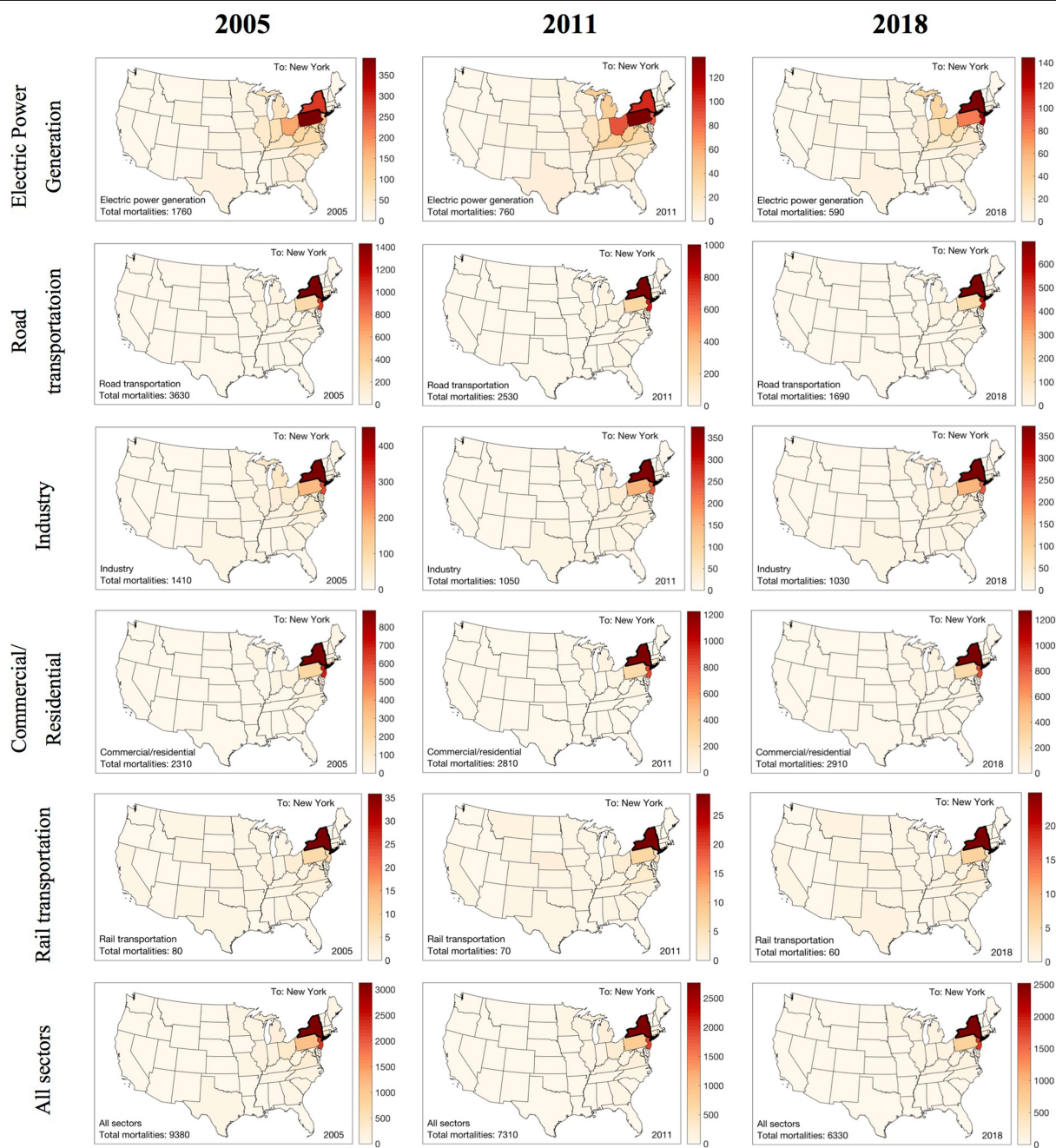
(ii) Emission species



Fraction of early deaths that cross state lines (%)

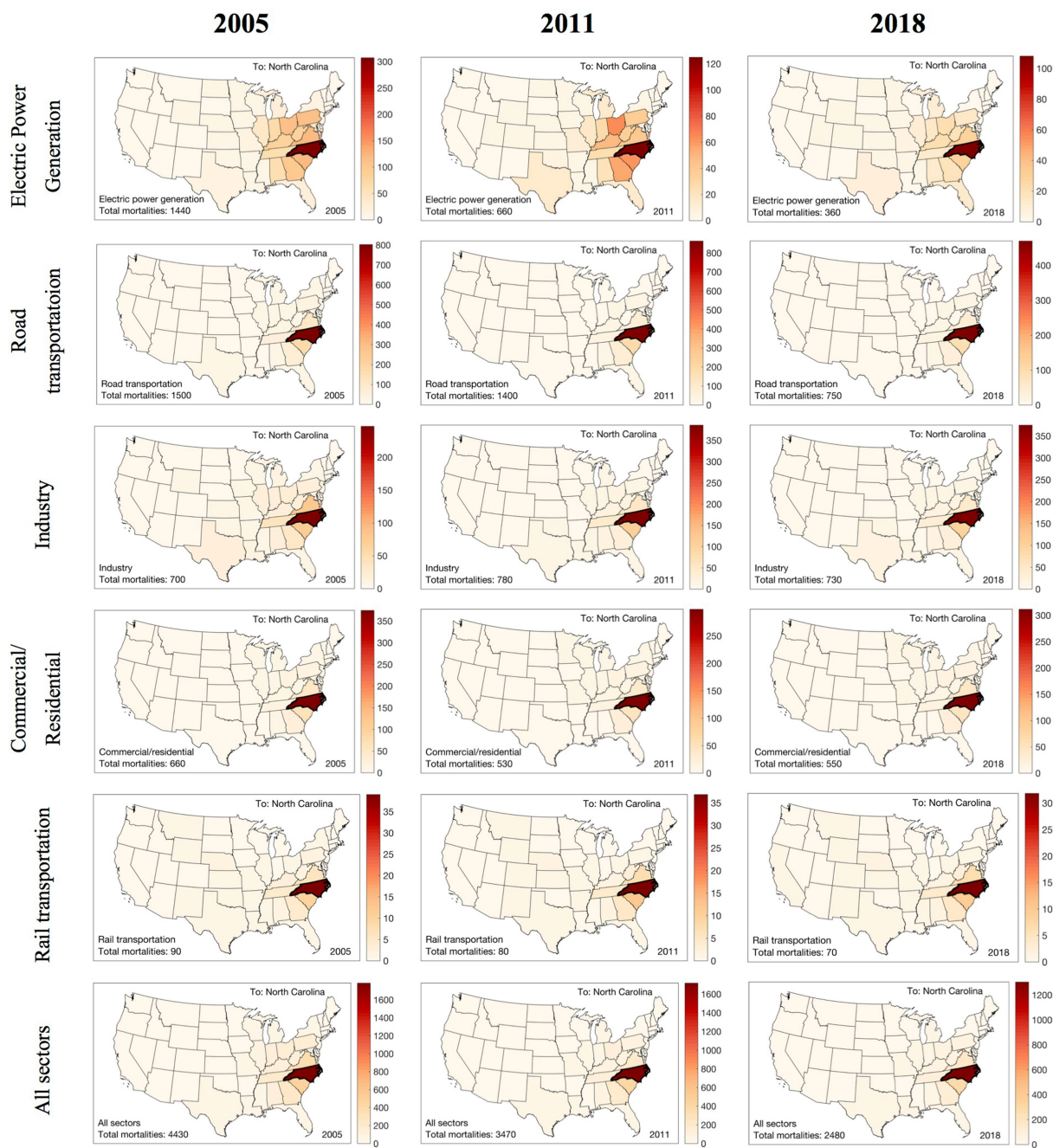
Extended Data Fig. 2 | Annual early-death source-receptor matrices for 2005, 2011 and 2018 for the contiguous US. Each matrix comprises 48 × 48 states. **a (i)**, The total source-receptor matrix for 2011. **a (ii)**, Its breakdown to PM_{2.5}-attributable and ozone-attributable impacts for all three years. **b**, Source-receptor early-death attribution to emission sectors (i) and emission species that lead to the formation of PM_{2.5} and/or ozone (ii). States are grouped in regions defined by the Bureau of Economic Analysis²⁰ (labelled in a) and

ordered from west (left) to east (right). The ordering of individual states is presented in Extended Data Fig. 1. Boxed percentages represent the fraction of impacts that occur out of the state that caused the corresponding emissions. We note that to obtain these summarized source-receptor matrices using conventional modelling approaches ('forward difference simulations') would have required around 1,300 simulations.



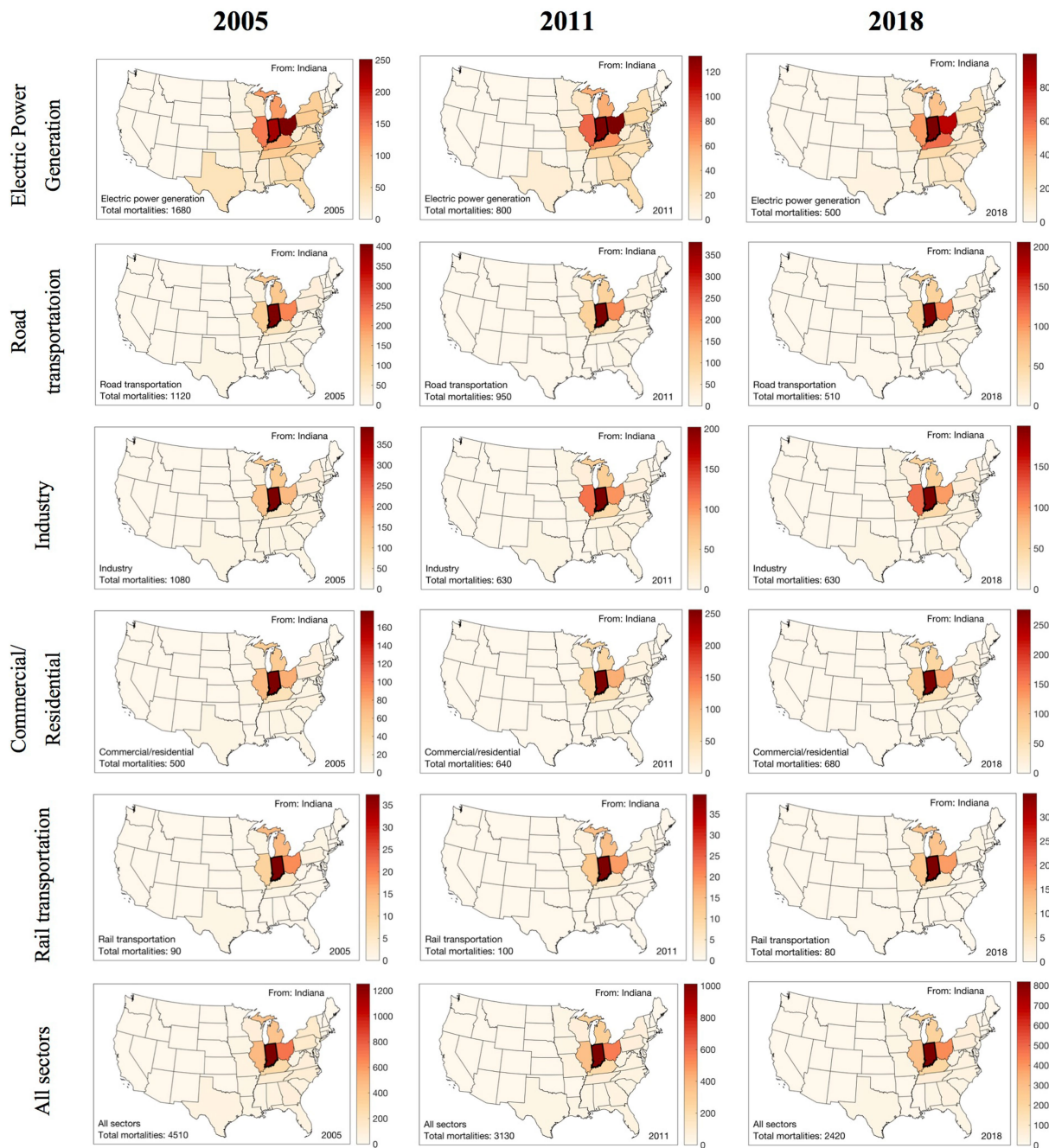
Extended Data Fig. 3 | Origins of New York annual early deaths, for 2005, 2011 and 2018, for five sectors and in total. Each state is coloured according to the annual early deaths that emissions from that state cause in the state of New

York, for each sector–year combination. The total early deaths occurring in New York (that is, the sum of all states' values) for each sector–year combination is displayed at the bottom left of each panel.



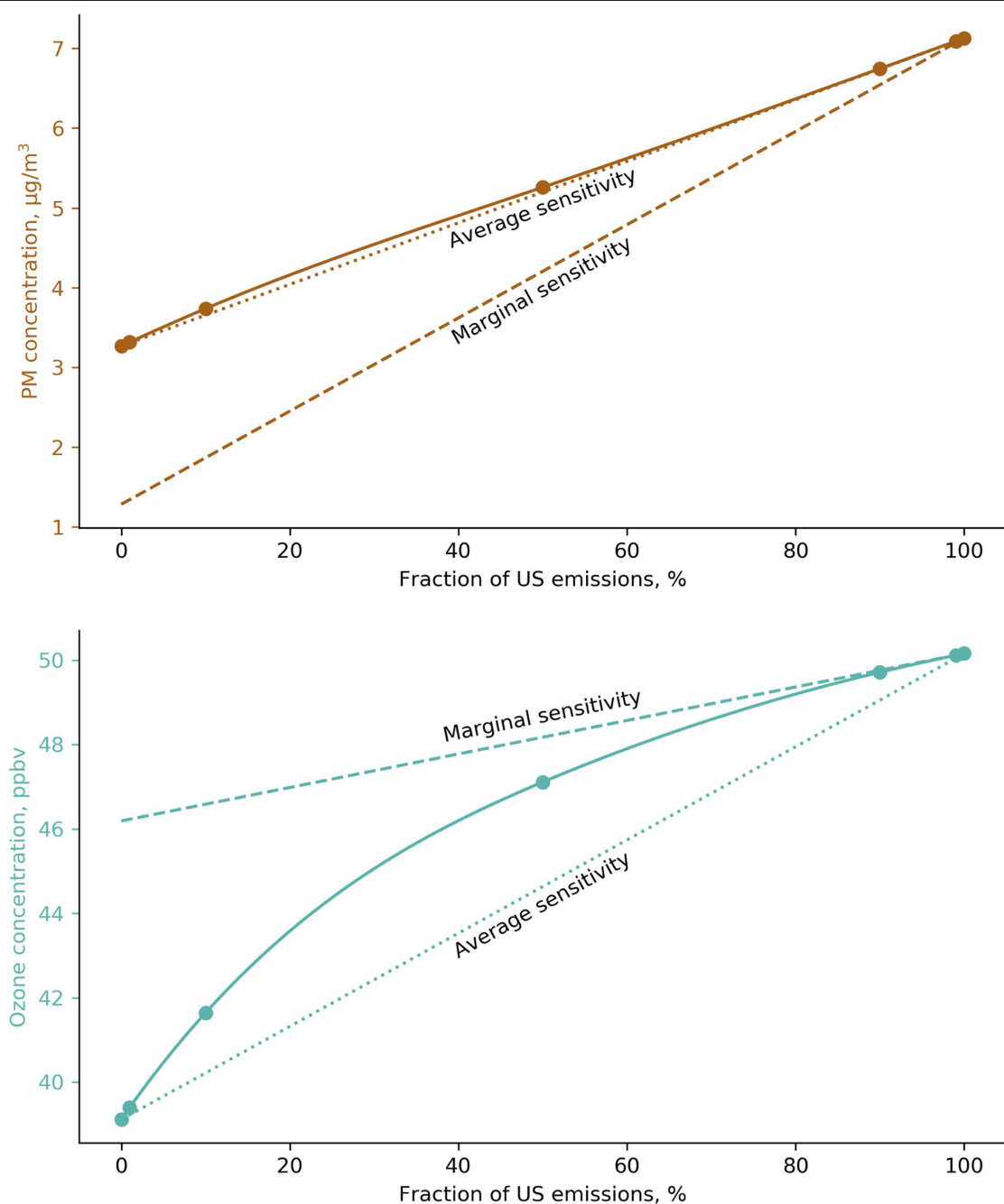
Extended Data Fig. 4 | Origins of North Carolina annual early deaths, for 2005, 2011 and 2018, for five sectors and in total. Each state is coloured according to the annual early deaths that emissions from that state cause in the

state of North Carolina, for each sector–year combination. The total early deaths occurring in North Carolina (the sum of all states’ values) for each sector–year combination is displayed at the bottom left of each panel.



Extended Data Fig. 5 | Receptors of annual early deaths due to emissions in Indiana for 2005, 2011 and 2018, for five sectors and in total. Each state is coloured according to the annual early deaths that occur in that state because

of emissions in Indiana, for each sector–year combination. The total early deaths caused by Indiana emissions (that is, the sum of all states’ values) for each sector–year combination is displayed at the bottom left of each panel.



Extended Data Fig. 6 | Changes in the response of surface-population-weighted $\text{PM}_{2.5}$ and ozone concentrations to US emissions. Data points show the results of a series of forward simulations, in which the input conditions of the simulation (the total US anthropogenic emissions of all species) are reduced, joined by a cubic spline fit. The 'average sensitivity' lines indicate the gradient implied when impacts due to all sectors combined are calculated—that is, when the effects of atmospheric nonlinearity are taken into account—

and thus the total results are scaled to match this. The 'marginal sensitivity' lines indicate the gradient of the response obtained by our GEOS-Chem adjoint simulation, and are used for calculations of individual sector and species impacts (where individual perturbations are of smaller size). The difference between the zero intercept of the two lines constitutes the 'interaction' effect. All values are population-weighted means for 2011.

Extended Data Table 1 | Primary PM_{2.5}, NO_x and SO_x emissions totals for 2005, 2011 and 2018

	2005				
	PM _{2.5}	NO _x	SO _x	NH ₃	CO
Electric power generation	0.46	3.42	9.46	0.02	0.57
Industry	0.57	2.75	2.55	0.13	3.03
Commercial/ Residential	0.69	0.76	0.49	0.04	4.82
Road transportation	0.27	8.17	0.16	0.14	39.3
Marine	0.07	1.30	0.45	0.00	0.18
Rail	0.03	1.01	0.07	0.00	0.11
Aviation (LTO %)	0.003 (15%)	0.60 (13%)	0.06 (15%)	--	0.26 (44%)
	2011				
	PM _{2.5}	NO _x	SO _x	NH ₃	CO
Electric power generation	0.18	1.78	4.10	0.02	0.69
Industry	0.41	2.52	1.20	0.09	2.49
Commercial/ Residential	0.62	0.61	0.22	0.12	3.67
Road transportation	0.19	5.10	0.03	0.11	23.1
Marine	0.02	0.60	0.05	0.00	0.11
Rail	0.02	0.77	0.01	0.00	0.12
Aviation (LTO %)	0.003 (13%)	0.58 (13%)	0.05 (13%)	--	0.17 (40%)
	2018				
	PM _{2.5}	NO _x	SO _x	NH ₃	CO
Electric power generation	0.19	1.42	1.32	0.04	0.74
Industry	0.52	2.15	0.85	0.09	2.42
Commercial/ Residential	0.64	0.60	0.12	0.12	3.82
Road transportation	0.11	2.38	0.01	0.08	14.0
Marine	0.01	0.37	0.002	0.00	0.09
Rail	0.02	0.67	0.001	0.00	0.14
Aviation (LTO %)	0.003 (12%)	0.63 (12%)	0.05 (12%)	--	0.15 (33%)

Emissions expressed in teragrams per year for each sector for 2005, 2011 and 2018. Emissions for all sectors apart from aviation are derived from EPA's NEI. Aviation emissions are taken from the AEDT inventory²⁹ (for years 2006, 2010/2012 and 2015) and include emissions that occurred over the contiguous US. The percentages of aviation emissions that occur within around 1 km of altitude (landing and take-off emissions) are given in parentheses, and are the aviations emissions included in our analysis.

Extended Data Table 2 | Five states with the greatest reduction in annual early deaths between 2005 and 2018

By each state				In each state			
State	2005	2018	$\Delta(2005-18)$ (%)	State	2005	2018	$\Delta(2005-18)$ (%)
WV	1,740 [1,240-2,250]	745 [540-950]	1,000 (57%)	AL	2,080 [1,520-2,640]	990 [730-1,260]	1,080 (53%)
AL	2,640 [1,920-3,350]	1,280 [950-1,610]	1,350 (51%)	ME	330 [240-430]	170 [120-220]	160 (48%)
TN	2,370 [1,730-3,020]	1,210 [880-1,550]	1,160 (49%)	MS	970 [710-1,220]	510 [380-640]	460 (47%)
MD	3,240 [2,250-4,230]	1,660 [1,140-2,180]	1,580 (49%)	VA	3,760 [2,680-4,840]	2,010 [1,440-2,590]	1,750 (46%)
KY	2,750 [1,940-3,560]	1,430 [1,020-1,850]	1,320 (48%)	WV	790 [570-1,010]	420 [310-540]	370 (46%)

Data are given in terms of early deaths caused by emissions from each state and in each state. Values in square brackets show 95% confidence intervals.

Extended Data Table 3 | Early deaths attributable to each sector and species (that lead to PM_{2.5} and/or ozone formation) for 2005, 2011 and 2018**(a)**

Sector	2005	2011	2018
Electric power generation	24,400 [16,900-31,800]	12,800 [9,100-16,600]	8,500 [6,000-10,900]
Industry	22,400 [15,600-29,100]	19,100 [13,400-24,700]	18,200 [12,900-23,600]
Commercial/residential	20,400 [14,100-26,800]	26,800 [18,300-35,400]	28,200 [19,200-37,300]
Road transportation	37,000 [26,500-47,600]	30,800 [21,800-39,800]	18,400 [12,900-23,800]
Marine Transportation	4,600 [3,200-5,900]	1,500 [1,100-2,000]	810 [590-1,000]
Rail transportation	2,300 [1,600-2,900]	2,400 [1,700-3,100]	2,100 [1,500-2,800]
Aviation	130 [80-200]	210 [140-270]	220 [150-290]
<i>Non-linear PM_{2.5} interactions</i>	<i>-29,500</i>	<i>-28,400</i>	<i>-23,500</i>
<i>Non-linear ozone interactions</i>	<i>+17,900</i>	<i>+18,000</i>	<i>+13,100</i>
Total	96,600 [74,200-125,000]	83,300 [62,400-104,200]	66,100 [49,300-82,900]

(b)

Species	2005	2011	2018
NO _x	30,000 [21,300-38,700]	28,600 [20,500-36,800]	19,600 [14,000-25,200]
SO ₂	21,000 [14,100-28,000]	9,900 [6,600-13,100]	4,300 [2,900-5,800]
Primary PM _{2.5}	34,800 [23,200-46,400]	31,000 [20,600-41,300]	30,200 [20,100-40,200]
NH ₃	14,400 [9,600-19,200]	16,700 [11,800-23,500]	17,400 [11,600-23,200]
Other	10,900 [8,000-13,800]	6,500 [4,700-8,300]	4,800 [3,500-6,200]
<i>Non-linear PM_{2.5} interactions</i>	<i>-29,500</i>	<i>-28,400</i>	<i>-23,500</i>
<i>Non-linear ozone interactions</i>	<i>+17,900</i>	<i>+18,000</i>	<i>+13,100</i>
Total	96,600 [74,200-125,000]	83,300 [62,400-104,200]	66,100 [49,300-82,900]

Values in square brackets are 95% confidence intervals. Only the mean effect is reported for the nonlinear interaction terms.

Extended Data Table 4 | Alternative CRF application to 2011 early deaths for all sectors, for PM_{2.5} and ozone

Pollutant	Study	Mortality end-points	Early deaths
PM _{2.5}	<i>Ref. 7</i>	<i>All-cause</i>	<i>55,200</i> <i>[36,800-73,600]</i>
	Ref. 2	All-cause	94,300 [37,700-169,700]
	Ref. 50	All-cause	66,800 [64,900-68,600]
	Ref. 9	All-cause	55,200 [36,800-73,600]
	Ref. 49	All-cause	124,200 [62,100-195,100]
	Ref. 51	All-cause	55,200 [9,200-101,200]
	Ref. 12	NCD+LRI*	75,000 [65,500-85,300]**
	Ref. 7	Cardiopulmonary	38,600 [29,700-48,200]
	Ref. 9	Cardiovascular	31,200 [14,200-45,400]
	Ref. 49	Cardiovascular	69,100 [37,200-106,300]
	Ref. 51	Cardiovascular	87,500 [54,000-123,500]
Ozone***	<i>Ref. 52</i>	<i>Respiratory</i> <i>(MDA8, annual avg.)</i>	<i>28,100</i> <i>[18,700-37,400]</i>
	<i>Ref. 52</i>	All-cause (MDA8, annual avg.)	59,500 [29,800-119,100]
	Ref. 50	All-cause (24-hr avg. warm season)	32,900 [29,900-35,900]
	Ref. 53	Respiratory (MDA1, warm season)	9,700 [2,400-16,300]

Atmospheric nonlinearity is taken into account. The CRFs used to calculate the estimates in the main text are shown in italics. As in the main text, we apply these to the 30-plus population, using corresponding data for disease-specific baseline incidence rates from the WHO for 2012. Uncertainty intervals (in square brackets) reflect the 95% confidence intervals for each CRF.

*The GEMM model health end-point is all nonaccidental deaths, almost all of which are due to noncommunicable diseases (NCDs) and lower respiratory infections (LRIs). We use the all-cause mortality incidence rate from the WHO, excluding all injury-related deaths.

**To estimate the early deaths from the GEMM model, we use the parameters provided in ref.12 for more than >25 years, excluding the Chinese male cohort study, and use the mean population-weighted concentration of PM_{2.5} in the US to determine the local relative risk per unit increase in exposure. The uncertainty intervals here reflect one standard error in parameter θ of the model.

***Note that the different CRF studies compared here assume different measures of ozone exposure (annual mean 8-hour maximum in ref. ⁵²; warm-season (April–September) mean in ref. ⁵⁰; and warm-season 1-hour daily maximum in ref. ⁵³). We apply all of these using the annual mean 8-hour maximum ozone exposure. MDA8, maximum daily 8-hour average.

Tobacco smoking and somatic mutations in human bronchial epithelium

<https://doi.org/10.1038/s41586-020-1961-1>

Received: 7 June 2019

Accepted: 29 December 2019

Published online: 29 January 2020

Kenichi Yoshida^{1,7}, Kate H. C. Gowers^{2,7}, Henry Lee-Six¹, Deepak P. Chandrasekharan², Tim Coorens¹, Elizabeth F. Maughan², Kathryn Beal¹, Andrew Menzies¹, Fraser R. Millar², Elizabeth Anderson¹, Sarah E. Clarke², Adam Pennycuik², Ricky M. Thakrar^{2,3}, Colin R. Butler^{2,3}, Nobuyuki Kakiuchi⁴, Tomonori Hirano⁴, Robert E. Hynds^{2,5}, Michael R. Stratton¹, Iñigo Martincorena¹, Sam M. Janes^{2,3,8*} & Peter J. Campbell^{1,6,8*}

Tobacco smoking causes lung cancer^{1–3}, a process that is driven by more than 60 carcinogens in cigarette smoke that directly damage and mutate DNA^{4,5}. The profound effects of tobacco on the genome of lung cancer cells are well-documented^{6–10}, but equivalent data for normal bronchial cells are lacking. Here we sequenced whole genomes of 632 colonies derived from single bronchial epithelial cells across 16 subjects. Tobacco smoking was the major influence on mutational burden, typically adding from 1,000 to 10,000 mutations per cell; massively increasing the variance both within and between subjects; and generating several distinct mutational signatures of substitutions and of insertions and deletions. A population of cells in individuals with a history of smoking had mutational burdens that were equivalent to those expected for people who had never smoked: these cells had less damage from tobacco-specific mutational processes, were fourfold more frequent in ex-smokers than current smokers and had considerably longer telomeres than their more-mutated counterparts. Driver mutations increased in frequency with age, affecting 4–14% of cells in middle-aged subjects who had never smoked. In current smokers, at least 25% of cells carried driver mutations and 0–6% of cells had two or even three drivers. Thus, tobacco smoking increases mutational burden, cell-to-cell heterogeneity and driver mutations, but quitting promotes replenishment of the bronchial epithelium from mitotically quiescent cells that have avoided tobacco mutagenesis.

Lung cancer kills more people globally than any other cancer, and 80–90% of those deaths are attributable to tobacco exposure^{1,2}. Our model for how tobacco causes lung cancer emphasizes direct mutagenesis from the numerous (more than 60) carcinogens in cigarette smoke^{4,5}, combined with indirect effects such as inflammation, immune suppression and infection. As recognized first in the sequencing of the *TP53* gene⁵ and more recently in genome-wide sequencing of lung cancers^{6–10}, tobacco exposure leads to both an increase in somatic mutational burden and an altered spectrum of mutations. Clones of lung cancer cells from a smoker typically have tens of thousands of somatic mutations^{6,7,9}; of these, a small handful—probably fewer than 20—drive the biology of the tumour^{11–13}.

Epidemiological studies have quantified the relationships between lung cancer and duration of smoking, intensity of smoking, type of smoking and timing of smoking cessation^{1–3,14}. Interpreting these observations from population cohorts in terms of the molecular basis for tobacco carcinogenesis is challenging. Under a model in which lung cancer requires n driver mutations, an exposure that, say, increases mutation rates by k -fold should increase incidence by around k^n across

a range of growth patterns¹¹. However, in a paradox first noted in 1971¹⁵, the dose–response relationship between the number of cigarettes smoked per day and the risk of lung cancer is linear^{3,14}—that is, k^1 —or, at most, weakly quadratic¹⁶. The benefits of smoking cessation likewise do not fit into multistage models of cancer in a straightforward manner¹⁵. By stopping smoking in middle age or earlier, smokers avoid most of the risk of tobacco-associated lung cancer. This benefit begins to emerge almost immediately and accrues steadily with time². Of two people who smoked the same total number of cigarettes across their lifetime, why the person with longer duration of cessation should have a lower risk of lung cancer is difficult to explain if tobacco induces carcinogenesis exclusively by increasing the mutational burden.

Sequencing single-cell-derived colonies

We recruited 16 individuals to assess the landscape of somatic mutations in normal bronchial epithelium: 3 children, 4 individuals who had never smoked ('never-smokers'), 6 ex-smokers and 3 current smokers (Supplementary Table 1). For ethical reasons, samples could only be

¹Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK. ²Lungs For Living Research Centre, UCL Respiratory, University College London, London, UK. ³Department of Thoracic Medicine, University College London Hospital, London, UK. ⁴Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan. ⁵CRUK Lung Cancer Centre of Excellence, UCL Cancer Institute, University College London, London, UK. ⁶Stem Cell Institute, University of Cambridge, Cambridge, UK. ⁷These authors contributed equally: Kenichi Yoshida, Kate H. C. Gowers. ⁸These authors jointly supervised this work: Sam M. Janes, Peter J. Campbell. *e-mail: s.janes@ucl.ac.uk; pc8@sanger.ac.uk

obtained from subjects who underwent a bronchoscopy for clinical indications. The never-smokers and current smokers had a bronchoscopy to investigate changes that were eventually diagnosed as benign. Of the ex-smokers, two had had a previous cancer treated with curative intent, and five had a carcinoma in situ or invasive squamous cell carcinoma that was the indication for the bronchoscopy. The children in the cohort underwent a bronchoscopy for investigation or follow-up of congenital anomalies: all had normal bronchial epithelium.

Samples of airway epithelium were obtained from biopsies or brushings of main or secondary bronchi. These were dissociated into single cells, and epithelial cells positive for epithelial cellular adhesion molecule (EpCAM⁺) were flow-sorted (one to a well) onto mouse feeder cells allowing basal cell attachment and growth (Extended Data Fig. 1a). Each cell was independently cultured to obtain single-cell-derived colonies that expressed the transcripts expected for basal cells of pseudostratified bronchial epithelium (Extended Data Fig. 1b). Around 15–40% of flow-sorted cells typically produced colonies (Extended Data Fig. 1c), confirming that the sequenced cells were drawn from a prevalent and representative population of epithelial cells. Colonies underwent whole-genome sequencing to an average coverage of 16× (Supplementary Table 2, Extended Data Fig. 2a, b). Using a xenograft pipeline to flag non-human sequencing reads, somatically acquired mutations were identified from reads specific to the human genome. In nearly all colonies, the variant allele fraction (VAF) of mutations was around 50% on average, which is consistent with contamination-free colonies derived from a single bronchial cell (Extended Data Fig. 2c). To remove variants that had possibly been acquired in vitro, we excluded mutations with a VAF of less than 30% that were present in only a single colony (Extended Data Fig. 2c). Occasional colonies had a low mean VAF (Extended Data Fig. 2d), consistent with seeding by two bronchial cells; these colonies were excluded from downstream analyses. We estimated that a sequencing depth of 8× gave a sensitivity for variants of 70–75%, and this increased to more than 95% at a depth of 15× (Extended Data Fig. 2e). The majority of colonies had a sequencing depth greater than 15×, and we set a minimum cut-off of 8× for inclusion.

The final dataset comprises catalogues of somatic mutations from the whole genomes of 632 single bronchial cells. Five patients had a squamous cell carcinoma or carcinoma in situ, three of which we also sequenced. Normal basal cells from these patients shared no clonal relationships with the carcinomas, and we found no systematic differences in mutational burden between normal cells in the vicinity of carcinoma in situ lesions and cells in regions that were histologically normal (Extended Data Fig. 2f).

Mutational burden

The burden of somatic substitutions per cell showed considerable heterogeneity both across the cohort and even within individual patients (Fig. 1a). Using linear mixed-effects (LME) models, we assessed factors that influenced the mutational burden (Supplementary Code). Single-base substitutions increased significantly with age, at an estimated rate of 22 per cell per year (95% confidence interval (CI), 20–25; $P = 10^{-8}$; Fig. 1b). Previous or current smoking significantly increased the mean burden of substitutions ($P = 0.0002$) by an estimated 2,330 per cell (95% CI, 1,180–3,480) in ex-smokers and 5,300 per cell (95% CI, 3,660–6,930) in current smokers.

The effects of age and smoking were expected but, more surprisingly, smoking also markedly increased the variability in mutational burden from cell to cell, even within the same individual. Among closely collocated cells from a small biopsy of normal airway from a given subject, the estimated standard deviation was 2,350 per cell for ex-smokers and 2,100 per cell for current smokers, compared with 140 per cell for children and 290 per cell for adult never-smokers ($P < 10^{-16}$ by LME for within-subject heterogeneity of variance across smoking categories). There was also heterogeneity between subjects: the estimated standard deviation in

mean substitution burden across individuals was 1,200 per cell for ex-smokers and 1,260 per cell for current smokers, compared to 90 per cell for non-smokers ($P = 10^{-8}$ by LME for heterogeneity of variance).

Although most cells in ex-smokers or current smokers had a considerably higher substitution burden than cells in never-smokers, a fraction of cells in these patients had burdens within the range expected for never-smokers of an equivalent age (Fig. 1c). For many of these patients, the distribution of mutational burden was distinctly bimodal, with one mode in the near-normal range and the other mode exhibiting a substantially increased mutational burden (Extended Data Fig. 3a). Notably, although cells with a near-normal mutational burden were rarely present in current smokers, their relative frequency was on average fourfold higher in ex-smokers (95% CI, 2.0–7.9-fold; $P = 3 \times 10^{-6}$ by log-linear model), typically accounting for 20–40% of all cells studied. Colonies with a near-normal mutational burden expressed the same set of airway basal cell genes as did colonies with an increased mutational burden, and had the same tightly associated, cobble architecture in culture (Extended Data Fig. 3b, c), confirming that they derived from bronchial epithelial cells.

Among current and ex-smokers, we found that mutational burden was not significantly correlated with the duration of cigarette smoking or the number of cigarettes smoked per day, even if near-normal cells were excluded. However, the small numbers of subjects and large within-subject heterogeneity limits our statistical power for this analysis, and definitive analysis will require much larger sample sizes.

Insertions and deletions (indels) showed similar associations as substitutions, increasing steadily with age (0.7 indels per cell per year; 95% CI, 0.6–0.8; $P = 10^{-6}$) and tobacco smoking (101 extra indels per cell in smokers; 51 in ex-smokers; $P = 0.001$; Extended Data Fig. 4a). Generally, the normal bronchial epithelial cells had few copy-number changes or structural variants (Extended Data Fig. 4b)—this represents a qualitative difference from lung cancers, which tend to have large numbers of structural abnormalities^{6,7,9,17}. There were occasional examples of more-complex structural events in the bronchial epithelial cells, including chromoplexy (Extended Data Fig. 4c) and even chromothripsis in a cell from a child (Extended Data Fig. 4d). The latter is particularly interesting, given that recent data suggest that driver-gene fusions in lung adenocarcinoma can arise through complex structural events that occur early in life¹⁷.

Mutational signatures

A range of mutational processes operate in lung cancers, driven both by the exogenous carcinogens present in tobacco smoke and by endogenous DNA damage. These processes leave characteristic signatures in the genome⁸. We built phylogenetic trees for each patient, and applied a Bayesian de novo mutational-signature discovery algorithm to mutations assigned to each branch. We also included samples from squamous cell lung cancers¹⁸ and control samples cultured in vitro¹⁹ in the signature analysis to maintain comparability with previous analyses⁸ (Fig. 2). Few mutations in our samples (typically 10–30 per cell) were attributed to SBS-18, the signature that accounted for all variants in the control samples¹⁹, which confirmed that mutations acquired in vitro were minimal in our dataset. Similar results emerged using a different mutational-signature algorithm²⁰ (Extended Data Fig. 5a–c).

A large proportion of mutations in all subjects was attributed to the endogenous mutational signature SBS-5, which accumulated linearly with age (Fig. 2c, d). As reported previously^{7,8}, the absolute number of mutations attributed to this signature was higher in those with a smoking history (ex-smokers 1,140 per cell, 95% CI, 590–1,700; current smokers 2,200 per cell, 95% CI, 1,590–2,810; $P < 10^{-16}$). Signature SBS-1—which comprises C>T mutations at CpG dinucleotides—contributed larger proportions of mutations in children than adults, but the absolute numbers of SBS-1-attributed mutations continued to increase linearly with age through adulthood (Fig. 2c, d). Presumably, then,

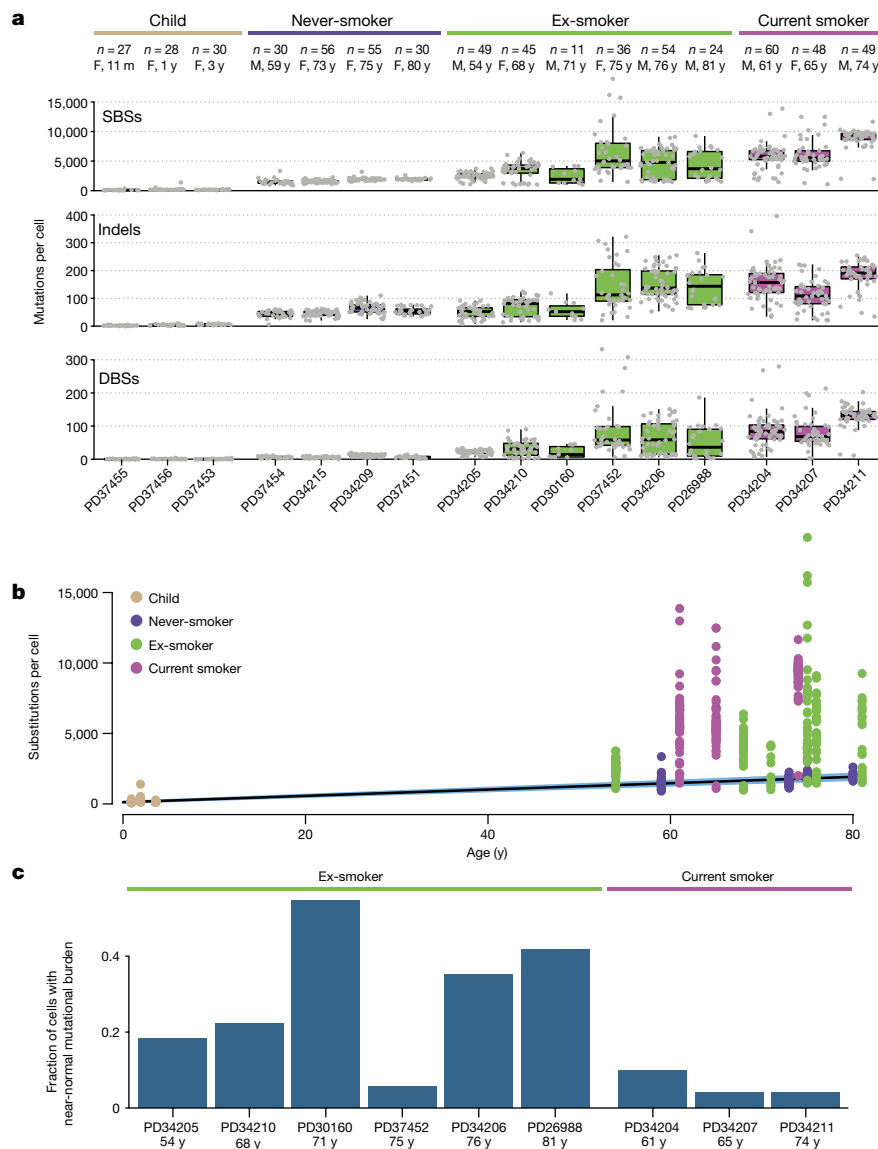


Fig. 1 | Mutational burden in normal bronchial epithelium. a, Burden of single-base substitutions (SBSs), small indels and double-base substitutions (DBSs) across patients in the cohort. The box-and-whisker plots show each subject, with the boxes indicating median and interquartile range and the whiskers denoting the range. The overlaid points are the observed mutational burden of individual colonies. **b**, Relationship of burden of substitutions per

cell with age. The points represent individual colonies ($n=632$) and are coloured by smoking status. The black line represents the fitted effect of age on the burden of substitutions, which was estimated from LME models after correction for smoking status and within-patient correlation structure. The blue shaded area represents the 95% CI for the fitted line. **c**, Fraction of cells with a near-normal mutational burden in current and ex-smokers.

SBS-1 is enriched during early lung development and continues steadily throughout life, but other signatures become proportionally more active in adulthood. A novel signature (Sig-A; Fig. 2b) was universally present across samples. It has some resemblance to SBS-5, and likewise increased linearly with age.

Signatures SBS-2 and SBS-13, which are caused by mutagenesis mediated by APOBEC3A or APOBEC3B, showed striking heterogeneity: they were mostly absent from bronchial cells, but occasionally contributed hundreds of mutations in an individual cell, even in children. This activity appeared to be temporally restricted: individual branches of a phylogenetic tree had high proportions of SBS-2 or SBS-13 despite their absence from antecedent and descendent branches (Fig. 3a, Extended Data Fig. 6). This implies that the episodic activity of APOBEC-mediated mutagenesis observed in cell lines²¹ extends to somatic cells in vivo, as the proportion of mutations attributed to APOBEC enzymes on a given branch of the phylogenetic tree does not predict past or future rates of mutagenesis in that lineage.

Three substitution signatures were largely restricted to current or ex-smokers. Signature SBS-4 was expected—this is the predominant signature in lung cancers from smokers^{7,8} and is recapitulated by in vitro exposure to polycyclic aromatic hydrocarbons¹⁹. SBS-16 comprised 5–15% of mutations in several current or ex-smokers, but was absent from never-smokers. This signature, with its distinctive pattern of transcription-coupled damage and repair²² (Extended Data Fig. 5d), correlates with alcohol and tobacco exposure in hepatocellular carcinomas^{8,23}, but has not been linked with tobacco exposure in lung cancers previously.

A new mutational signature (Sig-B) was extracted, which comprised predominantly T>A and T>C mutations and was evident only in patients with a history of smoking (Fig. 2b). The signature was mostly present at low rates, but in one ex-smoker it contributed up to 15% of mutations per cell. We found a strong transcriptional strand bias, whereby the transcribed strand showed decreased rates of mutation at the adenine in the

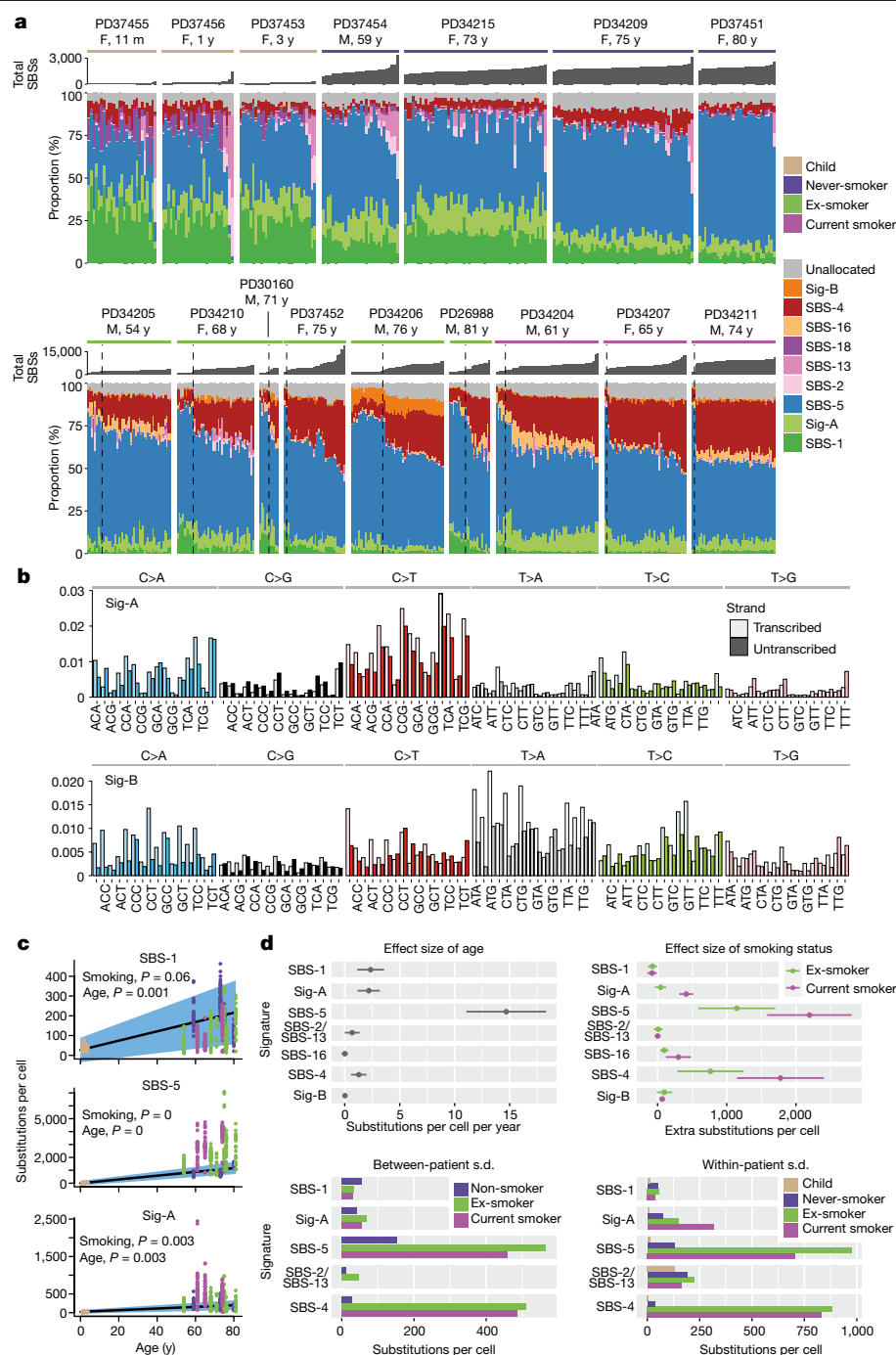


Fig. 2 | Mutational signatures in normal bronchial epithelium. a, Stacked bar plot showing the proportional contribution of mutational signatures to single-base substitutions across the $n = 632$ colonies from normal bronchial cells, extracted using a hierarchical Dirichlet process (HDP). Within each patient, colonies are sorted from left to right by increasing mutational burden (bar chart in dark grey above coloured signature-attribution stacks). The dashed black vertical lines in current and ex-smokers denote the cut-off between cells with a near-normal and an increased mutational burden. **b**, Trinucleotide context spectrum on transcribed and untranscribed strands of two new SBS signatures (Sig-A and Sig-B). The six substitution types are shown across the top. Within each substitution type, the trinucleotide context is shown as four sets of eight bars, grouped by whether an A, C, G or T, respectively, is 5' to the mutated base,

and within each group of eight by whether A, C, G or T is 3' to the mutated base. The activity of the mutational signature on the untranscribed strand is shown in a pale colour; on the transcribed strand it is shown in a darker colour. **c**, Number of base substitutions attributed to the three endogenous signatures across the cohort (y axis; $n = 632$ colonies) shown according to the age of the subject (x axis). The black line represents the fitted effect of age, which was estimated from LME models after correction for smoking status and within-patient correlation structure. The blue shaded area represents the 95% CI for the fitted line. The quoted P values for the fixed effects of age and smoking are derived from the full LME models. **d**, Estimated effect sizes of age, smoking status, between-patient and within-patient standard deviation of seven signatures (points) with 95% CIs (horizontal lines). Estimates are derived from LME models ($n = 632$).

T:A pairing. This is consistent with in vitro data that show that purines are more reactive than pyrimidines with mutagens in tobacco smoke⁵.

As described above, an unexpectedly high fraction of cells in ex-smokers had a near-normal mutational burden. These cells had considerably

lower proportions of SBS-4 mutations than cells with an increased mutational burden in the same patients. Instead, the distribution of signatures in these near-normal cells resembled that seen in never-smokers, with prominent endogenous signatures such as SBS-5, SBS-1

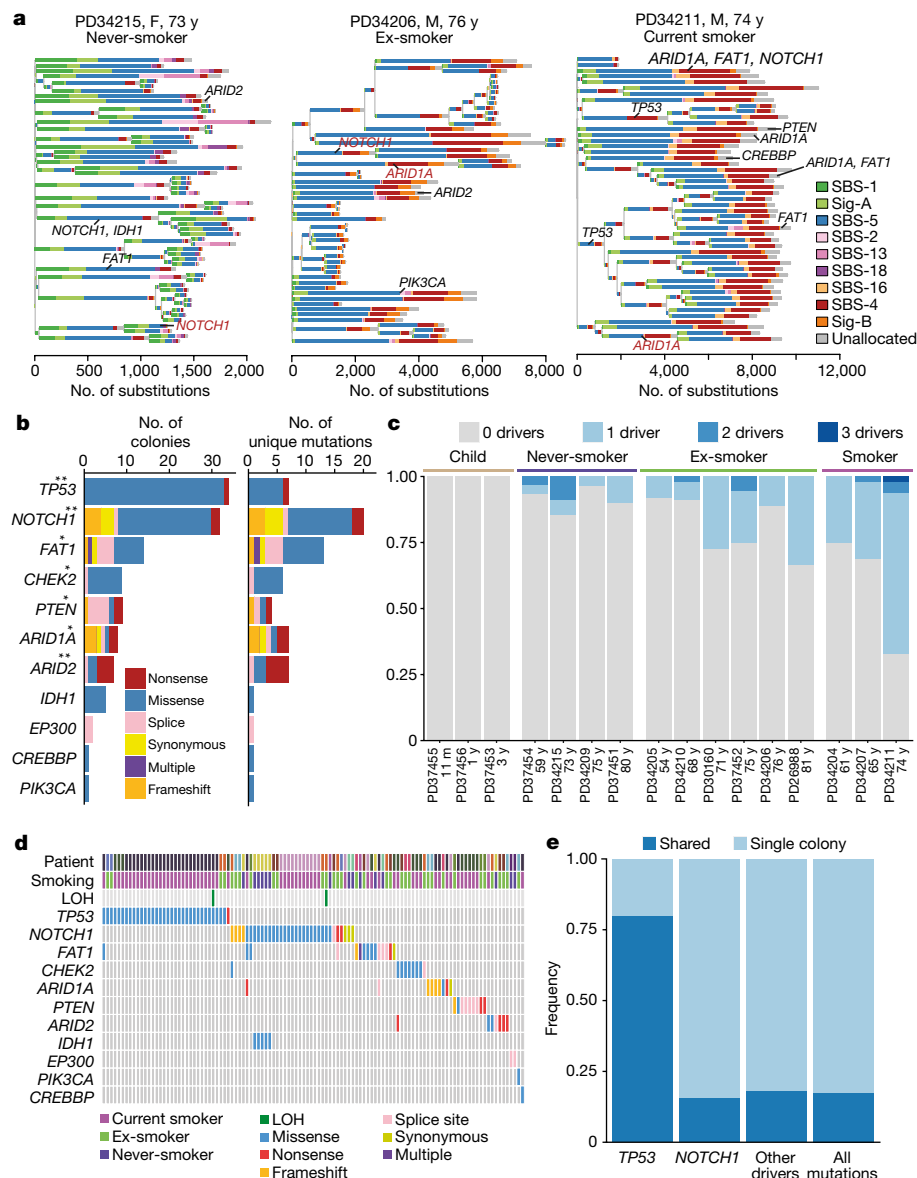


Fig. 3 | Driver mutations in normal bronchial epithelial cells. a, Phylogenetic trees showing clonal relationships among normal bronchial cells in three representative subjects. Branch lengths are proportional to the number of mutations (x-axis) specific to that clone or subclone. Each branch is coloured by the proportion of mutations on that branch that are attributed to the various SBS signatures. The driver mutations that were identified in each branch are also shown (black, SBS; red, indel). **b**, Total number of colonies with mutations (left) and number of unique mutations (right) in key cancer genes across the sample set ($n = 632$). ** represents genes that are significant ($q < 0.05$ by dNdScv) when correction for multiple-hypothesis testing is applied across all

coding genes; * represents genes that are significant ($q < 0.05$ by dNdScv) when correction for multiple-hypothesis testing is applied across known driver genes in lung cancers and normal squamous tissues (exact q values are provided in Supplementary Table 4). **c**, Fraction of colonies with 0, 1, 2 or 3 driver mutations across the 16 subjects. **d**, Distribution of driver mutations across colonies in the cohort, coloured by type of mutation. Loss of heterozygosity (LOH) that affects driver mutations is also shown. **e**, Frequency of driver mutations that are shared by more than one colony in a patient (dark blue) versus those found in a single colony (light blue) across different cancer genes.

and Sig-A. Phylogenetically, cells with a near-normal mutational burden showed polyclonal origins (Fig. 3a, Extended Data Fig. 6), suggesting that they do not arise from the expansion of a single ancestral cell.

Signatures of indels and double-base substitutions that were observed in normal bronchial epithelium matched those extracted from lung cancers²⁴ and those generated in vitro by exposure of cells to polycyclic aromatic hydrocarbons¹⁹ (Extended Data Figs. 7, 8). A history of tobacco smoking was particularly associated with a signature of double-base substitutions at CpC (equivalently GpG) dinucleotides—a finding that is in accordance with the high rates of C>A (G>T) single-base substitutions in SBS-4. Similarly, tobacco exposure was associated with an indel signature of single-base deletions of cytosines (guanines)

in our dataset. Together, these data suggest that the propensity of polycyclic aromatic hydrocarbons in tobacco smoke to bind guanine nucleotides can result in a range of mutation types even in normal bronchial epithelial cells, including single-base substitutions, dinucleotide substitutions and small indels.

Driver mutations

To assess whether any mutations are under positive selection in normal bronchial epithelium, we applied an algorithm, dNdScv, which identifies and quantifies the number of excess non-synonymous mutations compared with the number expected from the rate of synonymous

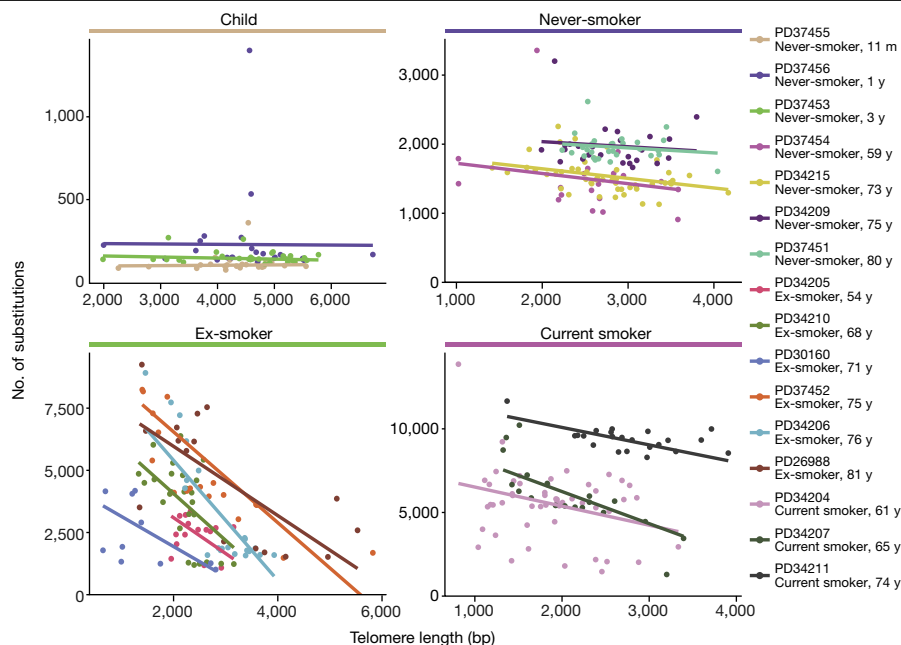


Fig. 4 | Relationship of telomere length with mutational burden. Split by smoking status, the graphs show the relationship between telomere length (x axis) and mutational burden (y axis) for colonies with less than 10% contamination from the mouse feeder cells ($n = 398$ colonies). Individual cells are shown as points and fitted lines for each patient are shown as coloured lines

(slopes were estimated using LME models). The difference in slopes according to smoking status is highly significant ($P = 0.0009$ for interaction term; LME models). One outlying cell from an ex-smoker, which had more than 10,000 mutations, was excluded from the plot to improve visualization.

(neutral) variants, correcting for local variation in mutation rates¹². With hypothesis testing applied across all coding genes, three were significant: *NOTCH1* (20 unique non-synonymous variants; $q = 1 \times 10^{-5}$); *TP53* (7 unique non-synonymous variants; $q = 2 \times 10^{-4}$); and *ARID2* (7 unique non-synonymous variants; $q = 4 \times 10^{-4}$) (Fig. 3b). When hypothesis testing was restricted to genes that are mutated in lung cancers^{12,13,18,25,26} and normal squamous tissues^{27–29}, *FAT1*, *PTEN*, *CHEK2* and *ARID1A* were also significant, showing the expected patterns of protein-truncating mutations (Supplementary Tables 3–5, Extended Data Fig. 9a). This set of significant genes closely resembles those under positive selection in squamous cell lung cancers^{13,18} and other normal squamous tissues^{27–30}.

Driver mutations were more frequent in patients with a history of tobacco smoking (Fig. 3c, Extended Data Fig. 9b). No candidate driver mutations were identified in cells from children, and 4–14% of cells in adult never-smokers had drivers; by contrast, in current smokers, at least 25% of cells carried at least one driver. Furthermore, a small fraction of cells in smokers had two or even three coding driver point mutations (Fig. 3d)—as many as is seen in some lung cancers¹². We used generalized LME models to quantify these effects (Supplementary Code). Driver mutations were significantly more frequent in individuals with a smoking history and showed an increase of 2.1-fold in current smokers compared to never-smokers (95% CI, 1.0–4.4; $P = 0.04$). The number of driver mutations also increased independently with age, with every decade of life increasing the number of drivers per cell by 1.5-fold (95% CI, 1.2–2.1; $P = 0.004$)—a pattern reminiscent of the increasing number of driver mutations with age in the oesophagus^{28,29}. Finally, the number of driver mutations doubled on average for every 5,000 extra somatic mutations per cell, independent of the other variables (95% CI, 1.4–2.7; $P = 0.0003$).

Layering driver mutations onto phylogenetic trees revealed that driver mutations occurred throughout molecular time (Fig. 3a, Extended Data Fig. 6). Mutations in *TP53* were much more likely to be shared by two or more sequenced cells (Fig. 3e), however, suggesting that they either occur earlier in molecular time or drive larger clonal expansions.

Telomere lengths

To assess historic mitotic activity, we estimated telomere lengths from the sequencing data (Fig. 4). Bronchial cells from children had longer telomeres than did cells from adults (Extended Data Fig. 10), as expected, and telomere length showed no correlation with mutational burden in children. Among never-smokers, there was also minimal correlation between mutational burden and telomere length. In current smokers, however—and especially in ex-smokers—there was a strong inverse relationship between telomere length and mutational burden, independent of the number of driver mutations ($P = 0.0009$ for interaction between smoking status and telomere length by LME models; Supplementary Code). In particular, the cells with a near-normal mutational burden in ex-smokers had considerably longer telomeres than did their more-mutated counterparts, suggesting that they have historically undergone fewer cell divisions.

Discussion

The simplicity of the notion that cigarette smoking causes lung cancer through its mutagenic effects belies the underlying complexity of how tobacco shapes clonal dynamics, mutation acquisition and the selective environment in the bronchus. As expected, exposure to tobacco smoke increases the number of somatic mutations (by an average of a few thousand mutations per normal bronchial cell); the excess mutations are attributable to signatures of carcinogens in cigarette smoke; and the increased mutational burden generates more driver mutations. What is unexpected, however, is the pronounced within-patient variation in mutational burden among smokers: cells from one small biopsy of bronchial epithelium can vary tenfold in their mutational burden, from 1,000 to over 10,000 mutations per cell.

Our cohort may be affected by recruitment bias, as samples could only ethically be obtained from individuals who underwent a clinically indicated bronchoscopy. Nonetheless, such a recruitment bias cannot explain the considerable within-patient variance in mutational

burden, and this finding probably therefore applies to smokers more generally. Understanding how heterogeneity in mutational burden among competing cells contributes to clonal evolution will be important for refining our models of lung cancer development, which usually assume that the effects of carcinogens are homogeneous across a population of cells. We recently described similar heterogeneity in tobacco-induced mutagenesis among neighbouring clones within non-malignant liver³¹, suggesting that this phenomenon is not restricted to bronchial epithelium.

We found that a qualitatively distinct population of bronchial epithelial cells with a near-normal mutational burden exists in subjects with a history of smoking. These cells have the same mutational burden as age-matched never-smokers; have low proportions of signatures from tobacco carcinogens and longer telomeres than more-mutated cells; and occur at a fourfold higher frequency in ex-smokers compared to current smokers. These cells are clearly protective against cancer—lung cancers that emerge in ex-smokers do not have a near-normal mutational burden, instead typically showing the high mutational burden that is associated with tobacco-induced signatures.

Two points remain unclear: how these cells have avoided the high rates of mutations that are exhibited by neighbouring cells, and why this particular population of cells expands after smoking cessation. The longer telomeres of these cells imply that cells with a near-normal burden have undergone fewer cell divisions, and therefore potentially represent recent descendants of quiescent stem cells. Although they remain elusive in human lung³², quiescent stem cells have been identified through lineage tracing in mouse models, and have been shown to occupy a protected niche in submucosal glands and expand after lung injury^{33–35}. A physically protected niche could explain how such stem cells avoid exposure to tobacco carcinogens, but so too could mitotic quiescence itself, as replication is required to convert adducted DNA bases to mutations.

It is tempting to assume that the expansion of cells with a near-normal burden after smoking cessation arises through better fitness in the altered selection landscape—perhaps because these cells have longer telomeres or fewer mutations, or because aberrant NOTCH or TP53 signalling confers less advantage in the absence of tobacco smoke. These explanations notwithstanding, the apparent expansion of the near-normal cells could represent the expected physiology of a two-compartment model in which relatively short-lived proliferative progenitors are slowly replenished from a pool of quiescent stem cells, but the progenitors are more exposed to tobacco carcinogens. Only in ex-smokers would the difference in mutagenic environment be sufficient to distinguish newly produced progenitors from long-term occupants of the bronchial epithelial surface.

Epidemiological studies show that the health benefits of stopping smoking begin immediately, accrue with time since cessation and are evident even after quitting late in life². That these benefits could be facilitated by replenishment of the bronchial epithelium with cells that are essentially impervious to decades of sustained cigarette smoking attests to the resilience and regenerative capacity of the lungs. The message for public health is that stopping smoking—at any age—does not just slow the accumulation of further damage, but can also reawaken cells that have not been damaged by past lifestyle choices.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1961-1>.

- Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. & Spivack, S. D. Epidemiology of lung cancer. Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* **143**, e15–e29S (2013).
- Peto, R. et al. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *Br. Med. J.* **321**, 323–329 (2000).
- International Agency for Research on Cancer. *Tobacco Smoke and Involuntary Smoking. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* Vol. **83** (IARC and World Health Organization, 2004).
- Hecht, S. S. Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol.* **21**, 160–171 (2008).
- Pfeifer, G. P. et al. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435–7451 (2002).
- Pleasant, E. D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
- Imielinski, M. et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).
- Alexandrov, L. B. et al. Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622 (2016).
- George, J. et al. Comprehensive genomic profiles of small cell lung cancer. *Nature* **524**, 47–53 (2015).
- Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
- Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G. & Vogelstein, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl Acad. Sci. USA* **112**, 118–123 (2015).
- Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e21 (2017).
- Campbell, J. D. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
- Garfinkel, L. & Stellman, S. D. Smoking and lung cancer in women: findings in a prospective study. *Cancer Res.* **48**, 6951–6955 (1988).
- Armitage, P. Response to Richard Doll: the age distribution of cancer. *J. Roy. Stat. Soc. A* **134**, 155–156 (1971).
- Doll, R. & Peto, R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J. Epidemiol. Community Health* **32**, 303–313 (1978).
- Lee, J. J.-K. et al. Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. *Cell* **177**, 1842–1857 (2019).
- Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
- Kucab, J. E. et al. A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836 (2019).
- Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. Mutational Patterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
- Petljak, M. et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *Cell* **176**, 1282–1294 (2019).
- Haradhvala, N. J. et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell* **164**, 538–549 (2016).
- Letouze, E. et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017).
- Alexandrov, L. et al. The repertoire of mutational signatures in human cancer. *Nature* <https://doi.org/10.1038/s41586-020-1943-3> (2020).
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
- Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
- Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
- Yokoyama, A. et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).
- Yizhak, K. et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, eaaw0726 (2019).
- Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).
- Teixeira, V. H. et al. Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *eLife* **2**, e00966 (2013).
- Hegab, A. E. et al. Novel stem/progenitor cell population from murine tracheal submucosal gland ducts with multipotent regenerative potential. *Stem Cells* **29**, 1283–1293 (2011).
- Tata, A. et al. Myoepithelial cells of submucosal glands can function as reserve stem cells to regenerate airways after injury. *Cell Stem Cell* **22**, 668–683 (2018).
- Lynch, T. J. et al. Submucosal gland myoepithelial cells are reserve stem cells that can regenerate mouse tracheal epithelium. *Cell Stem Cell* **22**, 653–667 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Subjects

Subjects were recruited at University College London Hospitals (UCLH) or Great Ormond Street Hospital (GOSH) and gave written informed consent with approval of the Research Ethics Committee (REC reference 06/Q0505/12 and 11/LO/152, respectively). Details of the patients studied are listed in Supplementary Table 1. All patients underwent bronchoscopy as part of their clinical care. In adults, the bronchoscopy procedure was performed for diagnostic or surveillance indications; in children, it was undertaken for investigational procedures on congenital tracheal abnormalities. For five patients with squamous cell carcinomas or carcinoma in situ, biopsy of normal bronchial tissue was taken from a site distant from the tumour.

Single-cell-derived colonies

Endobronchial biopsies were dissociated using 16 U/ml dispase in RPMI for 20 min at room temperature. The epithelium was dissected away from the underlying stroma and fetal bovine serum (FBS) was added to a final concentration of 10%. Both the epithelium and stroma were combined and digested in 0.1% trypsin/EDTA at 37 °C for 30 min. The solution was neutralized with FBS to a final concentration of 10% and added to the neutralized dispase solution³⁶. Cells were passed through a 100- μ m cell strainer and stained in sorting buffer (1 \times PBS, 1% FBS, 25 mM HEPES and 1 mM EDTA) with anti-CD45-PE (BD Pharminogen 555483, 1:200), anti-CD31-PE (BD Pharminogen 555446, 1:200), anti-EPCAM-APC (Biolegend 324208, 1:50) antibodies and DAPI (1 μ g/ml). For endobronchial brushings, no dissociation was carried out and the cell suspension was passed through a 100- μ m cell strainer before staining.

Cells were single-cell sorted on the basis of their expression of CD45, CD31 and EpCAM, using a BD FACSAria Fusion. Each DAPI⁺CD45⁺CD31⁺EpCAM⁺ cell was sorted into 1 well of a 96-well plate, pre-coated with collagen I and mitotically inactivated 3T3-J2 feeder cells. Feeder cells were authenticated by whole-genome sequencing, and were screened for mycoplasma contamination by PCR. Cells were grown in fresh epithelial growth medium³⁷ (Dulbecco's modified Eagle medium (DMEM):F12 at a 3:1 ratio with penicillin–streptomycin, 5% FBS, 5 μ M Y-27632, 5 μ g/ml insulin, 25 ng/ml hydrocortisone, 0.125 ng/ml epidermal growth factor, 0.1 nM cholera toxin, 250 ng/ml amphotericin B and 10 μ g/ml gentamicin), which was supplemented for the first week of culture with epithelial growth medium that had been conditioned on growing epithelial cells and a final concentration of 10 μ M Y-27632. Epithelial cells were grown in 96-well plates for 2 weeks before being passaged into 24-well plates and then into T25 flasks. Epithelial cells were in culture for a total of about 25 days at 37 °C and 5% CO₂ with 3 changes of medium per week. When cells reached 70–80% confluence in T25 flasks, they were differentially trypsinized (making use of the greater sensitivity of feeder cells to trypsin compared with epithelial cells), generating a mostly pure population of epithelial cells. DNA was then extracted using the PureLink Genomic DNA Mini Kit (Invitrogen).

Whole-genome sequencing

Paired-end sequencing reads (150 bp) were generated using the Illumina HiSeq X-Ten platform for 662 samples from 16 patients. The target coverage was 15 \times per sample, except for 30 \times for 26 pilot samples that were derived from the first patient (PD26988). For ten patients, blood DNA samples were also sequenced as germline controls. For three patients, samples of bulk squamous cell carcinoma or carcinoma in situ, which were collected at the same or close time points (around four months after), were sequenced, including two samples of carcinoma in situ

that were used in a previous study³⁸ (PD38326a and PD38327a, which are carcinomas in situ that were derived from PD30160 and PD34210, respectively). We also sequenced the whole genome of the pure mouse feeder cell layer.

Discrimination of human and mouse sequences

Bronchial epithelium samples were cultured on J2 mouse embryonic feeder fibroblast cells, which caused various degrees of contamination of mouse DNA in the samples from bronchial cell colonies. To remove mouse-derived sequencing reads, we used the Xenome algorithm³⁹ with default setting (*k*-mer size = 25). The Xenome algorithm classifies fastq files into five categories: graft (human), host (mouse), ambiguous, both and neither. We confirmed that most of the sequencing reads of a sample of pure human DNA were classified as human (98%) and those of a sample of DNA derived from mouse feeder cells were rarely (2.8%) classified as human (Extended Data Fig. 2a). In addition, we mapped sequencing reads of a DNA sample from mouse feeder fibroblasts to the human reference genome, and confirmed that most of the mouse-derived mutations had been successfully removed using Xenome for selected samples with mouse contamination (Extended Data Fig. 2b). Although all samples were negative for mycoplasma using standard laboratory PCR testing, Xenome identified sequencing reads derived from the mycoplasma genome in a subset of samples, and assigned them to the 'neither' classification.

With testing complete, we ran Xenome for all bronchial epithelium samples, and aligned only reads that were classified as human to the human reference genome (NCBI build 37d5) using the BWA-MEM algorithm. The metrics of sequencing coverage and proportion of human-derived reads are listed in Supplementary Table 2, and 20 samples with an average sequencing depth of less than 8 \times were excluded from further analysis owing to lower estimated sensitivity, as described later (Extended Data Fig. 2e).

Clonality of samples

To ensure that each sample was single-cell-derived, we visually inspected the distribution of VAFs of mutations: 632 clones had VAFs distributed around 50%, confirming that they were derived from a single cell, but 10 clones had lower allele fractions, suggesting that these colonies were oligoclonal (Extended Data Fig. 2d). These samples were removed from further analyses (Supplementary Table 2).

Single-base-substitution calling

Single-base substitutions were called using the Cancer Variants through Expectation Maximization (CaVEMan) algorithm⁴⁰ with copy-number options of major copy number 5, minor copy number 2 and normal contamination 0.1. To allow the discovery of early embryonic mutations, we ran CaVEMan using an unmatched normal control. In addition to the default 'PASS' filter, we removed variants with a median alignment score (ASMD) < 120 and those with a clipping index (CLPM) > 0, to remove mapping artefacts. Variants identified in the mouse feeder fibroblast DNA sample were also removed, if they persisted in the call-set. Subsequently, for every mutation identified in any colonies from each patient, we counted the number of mutant and wild-type reads in all bronchial samples from the same patient using the bam2R function of the R package deepSNV⁴¹, for which bases with ≥ 30 base quality and sequencing reads with ≥ 30 mapping quality were used. Further filters described below were applied to identify true somatic mutations and separate them from either germline variants or recurrent sequencing errors.

Removing germline variants (binomial filter)

We fitted a binomial distribution to the total variant counts and total depth at each single-base substitution site across all samples from one patient. To differentiate somatic variants from germline variants, we used a one-sided exact binomial test, with the null hypothesis that these variants were drawn from a binomial distribution with a success

Article

probability of 0.5 (0.95 for sex chromosomes in males). The alternative hypothesis was that these variants were drawn from distributions with lower success probabilities. Variants with P value $> 10^{-10}$ were considered as germline variants.

Removing errors (beta-binomial filter)

We fitted a beta-binomial distribution to the variant counts and depths of all single-base substitutions across samples from the same patient for the remaining somatic variants. The beta-binomial was used as it captures the difference between artefactual variant sites and true somatic variants. Many artefacts appear to be randomly distributed across samples and can be modelled as drawn from a binomial distribution. True somatic variants will be present at a high VAF in some samples, but absent in others, and are hence best captured by a highly overdispersed beta-binomial. For each variant site, the maximum likelihood of the overdispersion factor (p) was calculated using a grid-based method (ranging from a value of 10^{-6} to $10^{-0.05}$). Variants with $p > 0.1$ were filtered out and considered to be artefactual. The code for this filter is based on the Shearwater variant caller⁴¹.

Removing mutations that were induced in vitro

We observed peaks of lower VAFs in a subset of samples (Extended Data Fig. 2c), which suggested that mutations were present that had arisen during the in vitro expansion of the single cell. These peaks were more prominent in samples from children, suggesting that the number of this kind of mutation is relatively small. They would, however, be more prominent in samples with a low true mutational burden, such as in children. We discarded mutations with a median VAF ≤ 0.3 for autosomal regions and median VAF ≤ 0.6 for sex chromosomes across all samples from the same patient; these cut-offs were determined on the basis of the observed distribution of VAFs here and in a previous report²⁰.

We quantified sensitivity by measuring how well our algorithms called heterozygous germline polymorphisms in the colonies depending on sequencing depth; as our colonies are single-cell-derived, we would expect heterozygous germline single-nucleotide polymorphisms to have the same VAF distribution as true somatic mutations in that original single cell. We find that a sequencing depth of $8\times$ leads to an estimated sensitivity of 70–75%, increasing to more than 95% at a sequencing depth of $15\times$. The majority of the colonies that we sequenced had depths of greater than $15\times$, and we set a minimum cut-off depth of $8\times$ for inclusion of a colony within the study (Extended Data Fig. 2e). Finally, we visually inspected allelic counts of removed germline variants with two or more samples without any mutant reads, and rescued embryonic mutations. Somatic variants were annotated using ANNOVAR⁴².

Indel calling

Indels were called using cgppindel⁴³, and an unmatched normal sample was used as the germline control. Indels that were detected in mouse fibroblast feeder cells were removed as mouse-derived artefacts. For all indels, indel-positive or -negative sequencing reads were counted using cgpVAF across all samples from each patient.

To remove germline variants and recurrent sequencing errors, the same binomial and beta-binomial filters were used as described above for single-base substitutions. We discarded mutations with a median VAF ≤ 0.25 for autosomal regions and median VAF ≤ 0.5 for sex chromosomes across all samples from the same patient to remove mutations that were induced in vitro.

Double-base-substitution calling

We first identified candidate double-base substitutions based on side-by-side single-base substitutions that were called using CaVEMan for each patient, and ran cgpVAF across all samples from each patient to remove those called in independent reads. Double-base substitutions

with three or more mutant reads in at least one sample were considered as true positives. Germline variants, errors and mutations induced in vitro were filtered as for single-base substitutions and indels.

Structural-variant calling

Structural variants were called using the BRASS algorithm⁴⁴, and matched normal samples (including blood samples and normal bronchial samples that were assigned on distantly located branches in phylogenetic trees) were used as controls. To remove germline structural variants, we filtered structural variants that were detected in the descendant colonies of both of the earliest two branches at the top of the phylogenetic tree for each patient. If the earliest branch had three or more branches (polytomy), those detected in both descendant and non-descendant samples of the earliest branch with the highest number were removed. We further filtered structural variants that were not identified using an unmatched normal control, to remove structural variants that were not filtered owing to the lower sequencing coverage of the matched normal control sample. Structural variants that were detected in other patients were also removed as germline variants or errors. Finally, all remaining structural-variant calls were manually inspected using the Integrated Genomics Viewer (IGV) to confirm somatic variants.

Copy-number calling

Copy-number changes were called using the ASCAT algorithm^{45,46}, and the same matched normal control samples as those used in the structural-variant analysis were used as germline controls. Copy-number gains, losses and copy-neutral LOHs were visually confirmed using LogR and BAF plots in ascatNgs. For amplification, those copy-number changes that were greater than 100 kb were visually confirmed using ascatNgs and JBrowse⁴⁷.

Mutational burden and estimation of the effects of age and smoking

For single-base substitutions, indels and double-base substitutions, samples with three or more mutant reads and a VAF of 0.2 or higher were considered to be mutated, and the number of each class of genetic lesions was counted for all bronchial cells. For structural variants, chromoplexy⁴⁸ (Extended Data Fig. 4c), chromothripsis⁴⁹ (Extended Data Fig. 4d) and translocation pairs with similar breakpoints were considered as single structural variants. Genetic lesions that were identified both as structural variants and copy-number changes were also considered as single events.

An LME model was then fitted to estimate the effects of age and smoking status on the number of single-base substitutions or indels using the nlme package in R (Supplementary Code). In addition to the fixed effects of age and smoking, patient was used as a grouping variable in the random effect, in which smoking status was used as a modifier of between-patient difference. Difference of within-group heterogeneity (heteroscedasticity) according to smoking status was also fitted in this model. The intercept of this model was probably derived from embryonic mutations and mutations that were introduced in vitro. Models were fitted using maximum likelihood estimation, and nested models were compared using likelihood ratio tests.

Identification of near-normal lung cells

We define cells as having a near-normal mutational burden if they have a mutational burden that is less than two non-smoker within-patient standard deviations plus two non-smoker between-patient standard deviations above the estimated number of mutations accumulated at the age of that patient using an LME model (Supplementary Code). The fraction of cells with a near-normal mutational burden was compared between current smokers and ex-smokers with log-linear regression using the logarithm of the total number of cells sequenced per patient as an offset.

Construction of phylogenetic trees

Phylogenetic trees were built with maximum parsimony using substitutions for each patient. First, the input matrix of mutations was made, in which samples with a VAF of 0.2 or higher and three or more mutant reads were considered as mutated samples and labelled as '1', and remaining samples were labelled as '0'. Among samples labelled as 0, samples with (i) a sequencing depth of $6\times$ or less for each mutated base and (ii) one or more mutant reads were considered as undetermined and labelled as '?'. For every individual, phylogenetic trees were constructed using the Camin–Sokal method of the Mix program of the RPhylip package, and consensus trees of all the trees were then constructed using the Consensus program in RPhylip.

Subsequently, all mutations were reassigned to branches in the phylogenetic trees. If mutations were called in all the descendants of a given branch and in no samples that were not descendants of the branch, mutations were perfectly assigned to those branches. Given the existence of samples with relatively lower sequencing depths for each mutated position, we also assigned mutations to branches if mutations were called in all but one undetermined descendant labelled as ? of a given branch, and all samples that were not descendants of the branch were wild-type (0). Given the smaller number of indels and double-base substitutions, these were assigned to each branch using the tree defined from single-base substitutions, rather than generating new trees for the other mutation types.

Extraction of SBS signatures

To analyse mutational signatures for single-base substitutions, those assigned to each branch of the phylogenetic trees were categorized into 288 subtypes, consisting of 6 mutation classes by 16 5'- and 3'-base contexts on the transcribed strand, non-transcribed strand or intergenic region. Mutational signatures were extracted using the HDP package⁵⁰, relying on the hierarchical Bayesian Dirichlet process (<https://github.com/nicolaroberts/hdp>). Owing to the lack of reference signatures categorized into 288 subtypes, we conducted a de novo signature extraction. We included somatic mutations from squamous cell lung carcinomas sequenced by The Cancer Genome Atlas (TCGA) and from in vitro single-cell culture controls as separate samples to maintain comparability with signatures that have already been established in previous studies. For identified SBS signatures, signatures with ≥ 0.90 cosine similarity with reported signatures, in terms of distribution to 96 or 192 subtypes²⁴, were considered as the same signatures, including SBS-1, SBS-4, SBS-5, SBS-16 and SBS-18. For the remaining new signatures, the expectation-maximization algorithm was used to deconvolute these signatures into the five signatures above and other known signatures in lung cancers (SBS-2, SBS-8 and SBS-13), because it is difficult to separate signatures that are strongly correlated across samples. If a signature reconstituted from the components that expectation maximization extracted (only including signatures that accounted for at least 10% of mutations in each sample to avoid overfitting) had a ≥ 0.90 cosine similarity to the original HDP signature, the signature was presented as its expectation-maximization deconvolution. Two HDP signatures met these criteria: one new signature was deconvoluted into a mixture of SBS-4 and SBS-5, and another new signature was deconvoluted in SBS-2 and SBS-13. After these analyses, seven known and two new SBS signatures were identified.

To validate these signatures that were identified using the HDP, we also analysed SBS signatures using the MutationalPatterns package²⁰, which relies on non-negative matrix factorization. The optimal factorization rank (7) was determined on the basis of the slope of the cophenetic correlation coefficient. MutationalPatterns identified similar signatures to SBS-5 (Signature A), SBS-4 (Signature B), Sig-B (Signature D), SBS-18 (Signature E), SBS-1 (Signature F), SBS-2 and SBS-13 (Signature G) (Extended Data Fig. 5a, b).

Extraction of indel and DBS signatures

For indels and double-base substitutions, each type of genetic alteration that was assigned to each branch of the phylogenetic trees was categorized into 83 and 78 subtypes, as previously reported²⁴. First, the algorithm was conditioned on the set of mutational signatures that have been detected in lung cancers (ID-1, ID-2, ID-3, ID-5, ID-6, ID-8, ID-9, DBS-2, DBS-4, DBS-5, DBS-6, DBS-11). This allows simultaneous discovery of known and new signatures. For known signatures, signatures identified by HDP with a cosine similarity ≥ 0.90 with corresponding reported signatures were accepted as known signatures. Deconvolution of new signatures to the above known signatures was also performed, and one new indel signature was deconvoluted in ID-5 and ID-8. Finally, ten known signatures and one new signature were identified.

Analysis of A>G transcriptional strand bias

First, we measured the distance from mutations to the nearest transcription start sites (TSSs) of all the expressed genes in the lung; expressed genes were defined as those with a median of one or more transcripts per million in lung samples in the GTEx database (<https://gtexportal.org/home/>). Mutations in regions of bidirectional transcription were excluded from further analysis. We tiled 10 kb up and downstream of the TSSs into 1-kb bins, and counted the number of A>G mutations on transcribed and untranscribed regions in each tile. This number was further divided by the average number of bins in intergenic regions.

Analysis of driver variants

To systematically identify genes under positive selection in normal bronchial epithelium, we used the dN/dS method¹². We performed exome-wide dN/dS analysis and also analysed global dN/dS ratios for driver genes ($n = 86$) reported in lung cancer^{12,13,18,26} or normal skin or oesophagus tissues^{27–29} using dNdScv (Supplementary Table 3). Genes with q value ≤ 0.05 were reported as driver genes (Supplementary Tables 4, 5). Finally, hot-spot mutations reported in COSMIC for four or more patients were also considered as driver mutations, in addition to those in the seven driver genes identified by dNdScv (Fig. 3b). The proportion of shared mutations (found in more than one colony) and private mutations (found in a single colony) was calculated for patients other than PD30160 (who had a low number of sequenced samples ($n = 13$)). For known lung cancer driver genes, the distributions of mutations were compared between bronchial cells and lung squamous cell carcinoma¹³ (Extended Data Fig. 9b).

To estimate the effect of smoking status on the number of driver mutations, a generalized LME model was fitted using the lme4 package in R (Supplementary Code). Patient was modelled as a random effect, and the fixed effects of age, smoking status and total mutational burden were fitted into the model.

Estimation of telomere length

The average telomere lengths of bronchial epithelium cells were estimated from the whole-genome sequencing data using Telomerecat⁵¹. Considering the similarity of telomere sequences between human and mouse, we aligned all sequencing reads to the human reference genome using BWA-MEM without using Xenome, and then ran Telomerecat on the bam files. Samples with reported mouse contamination of more than 10% were excluded from further analysis to prevent a possible effect of mouse cells on telomere length. The average telomere length for the mouse fibroblast feeder samples was estimated at 1,745 bp, which is within the range of estimates of human telomere length, so a low level of mouse contamination will not substantially affect the estimates.

An LME model was then fitted to estimate the effect of telomere length on the number of single-base substitutions using the lme4 package in R (Supplementary Code). Patient was modelled as a random effect, and the fixed effects of telomere length and its interaction with

smoking status, as well as the fixed effects of age and smoking status, were fitted into the model.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequencing data have been deposited at the European Genome-phenome Archive (<http://www.ebi.ac.uk/ega/>) under the accession number EGAD00001005193. Somatic-mutation calls, including single-base substitutions, indels and structural variants, from all 632 samples have been deposited on Mendeley Data with the identifier: <https://doi.org/10.17632/b53h2kwpyy.2>.

Code availability

Detailed method and custom R scripts for the analysis of mutational burden in bronchial epithelium are available in Supplementary Code. Other packages used in the analysis are as follows: R v.3.5.1; BWA-MEM v.0.7.17-r1188 (<https://sourceforge.net/projects/bio-bwa/>); CaVEMan v.1.11.2 (<https://github.com/cancerit/CaVEMan>); Pindel v.2.2.5 (<https://github.com/cancerit/cgpPindel>); Brass v.6.1.2 (<https://github.com/cancerit/BRASS>); ASCAT NGS v. 4.1.2 (<https://github.com/cancerit/ascatNgs>); Xenome (<https://github.com/data61/gossamer/blob/master/docs/xenome.md>); deepSNV v.1.28.0 (<https://bioconductor.org/packages/release/bioc/html/deepSNV.html>); ANNOVAR (<http://wannovar.wglab.org/>); IGV (<http://software.broadinstitute.org/software/igv/>); JBrowse (<https://jbrowse.org/>); cgpVAF (<https://github.com/cancerit/vafCorrect>); RPhylip v.0.1.23 (<http://www.phytools.org/Rphylip/>); hdp v.0.1.5 (<https://github.com/nicolaroberts/hdp>); MutationalPatterns v.1.8.0 (<https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html>); dNdScv v.0.0.1 (<https://github.com/im3sanger/dndscv>); and Telomerecat v.3.1.2 (<https://github.com/jhrf/telomerecat>).

36. Gowers, K. H. C. et al. Optimized isolation and expansion of human airway epithelial basal cells from endobronchial biopsy samples. *J. Tissue Eng. Regen. Med.* **12**, e313–e317 (2018).
37. Butler, C. R. et al. Rapid expansion of human epithelial stem cells suitable for airway tissue engineering. *Am. J. Respir. Crit. Care Med.* **194**, 156–168 (2016).
38. Teixeira, V. H. et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nat. Med.* **25**, 517–525 (2019).
39. Conway, T. et al. Xenome—a tool for classifying reads from xenograft samples. *Bioinformatics* **28**, 172–178 (2012).
40. Jones, D. et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. *Curr. Protoc. Bioinformatics* **56**, 15.10.1–15.10.18 (2016).

41. Gerstung, M., Papaemmanuil, E. & Campbell, P. J. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics* **30**, 1198–1204 (2014).
42. Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* **10**, 1556–1566 (2015).
43. Raine, K. M. et al. cgpPindel: identifying somatically acquired insertion and deletion events from paired end sequencing. *Curr. Protoc. Bioinformatics* **52**, 15.7.1–15.7.12 (2015).
44. Campbell, P. J. et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).
45. Raine, K. M. et al. ascatNgs: identifying somatically acquired copy-number alterations from whole-genome sequencing data. *Curr. Protoc. Bioinformatics* **56**, 15.9.1–15.9.17 (2016).
46. Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).
47. Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
48. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
49. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
50. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* <https://doi.org/10.1038/s41586-019-1913-9> (2020).
51. Farmer, J. H. R. et al. Telomerecat: a ploidy-agnostic method for estimating telomere length from whole genome sequencing data. *Sci. Rep.* **8**, 1300 (2018).

Acknowledgements This work was supported by a Cancer Research UK Grand Challenge Award (C98/A24032) and the Wellcome Trust. P.J.C. and S.M.J. are Wellcome Trust Senior Clinical Fellows (WT088340MA); S.M.J. receives funding as a member of the UK Regenerative Medicine Platform (UKRMP2) Engineered Cell Environment Hub (MRC; MR/R015635/1) and the Longfords BREATH lung regeneration consortium, and is further supported by The Rosetrees Trust, the Stonegate Trust, the British Lung Foundation and the UCLH Charitable Foundation; K.Y. is supported by a Japan Society for the Promotion of Science (JSPS) Overseas Research Fellowship and The Mochida Memorial Foundation for Medical and Pharmaceutical Research; S.M.J. and R.E.H. are supported by the Roy Castle Lung Cancer Foundation; R.E.H. is a Wellcome Trust Sir Henry Wellcome Fellow (WT209199/Z/17/Z); and I.M. is funded by Cancer Research UK (C57387/A21777). The authors thank S. Broad and F. Watt for providing 3T3-J2 fibroblasts, and B. Carroll for help with sample collection.

Author contributions S.M.J., P.J.C., K.Y., K.H.C.G. and H.L.-S. designed the experiments. K.H.C.G. performed all of the sample collection, cell isolation, clonal expansion and DNA extraction, with help from D.P.C., E.F.M. and F.R.M. E.F.M. and C.R.B. collected the paediatric samples, and E.F.M., D.P.C. and R.M.T. collected the adult samples. E.A. made sequencing libraries. K.Y. performed most of the data curation and statistical analysis, with H.L.-S., T.C., K.B., A.M., N.K. and T.H. providing assistance and advice. S.E.C. oversaw all of the clinical data collection and curation, and performed the flow cytometry characterization of the clones. R.E.H. and K.H.C.G. performed the qPCR characterization of the clones. M.R.S. oversaw the analysis of mutational signatures. P.J.C. and I.M. oversaw statistical analyses. R.E.H., A.P., K.H.C.G., K.Y., S.M.J. and P.J.C. performed data interpretation and, together with D.P.C., helped to draft and revise the manuscript.

Competing interests The authors declare no competing interests.

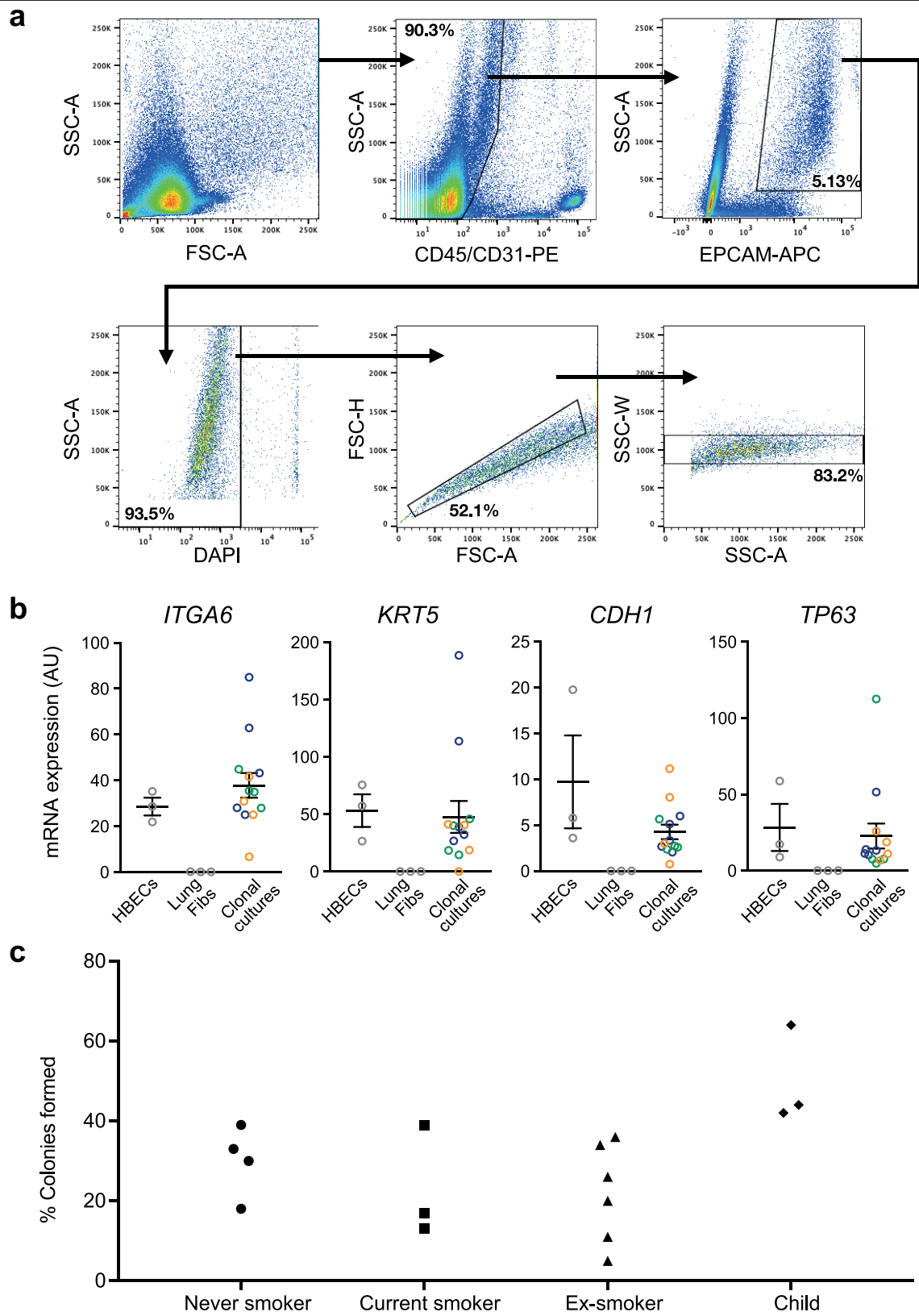
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1961-1>.

Correspondence and requests for materials should be addressed to S.M.J. or P.J.C.

Peer review information Nature thanks Gerd P. Pfeifer, Roman Thomas and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

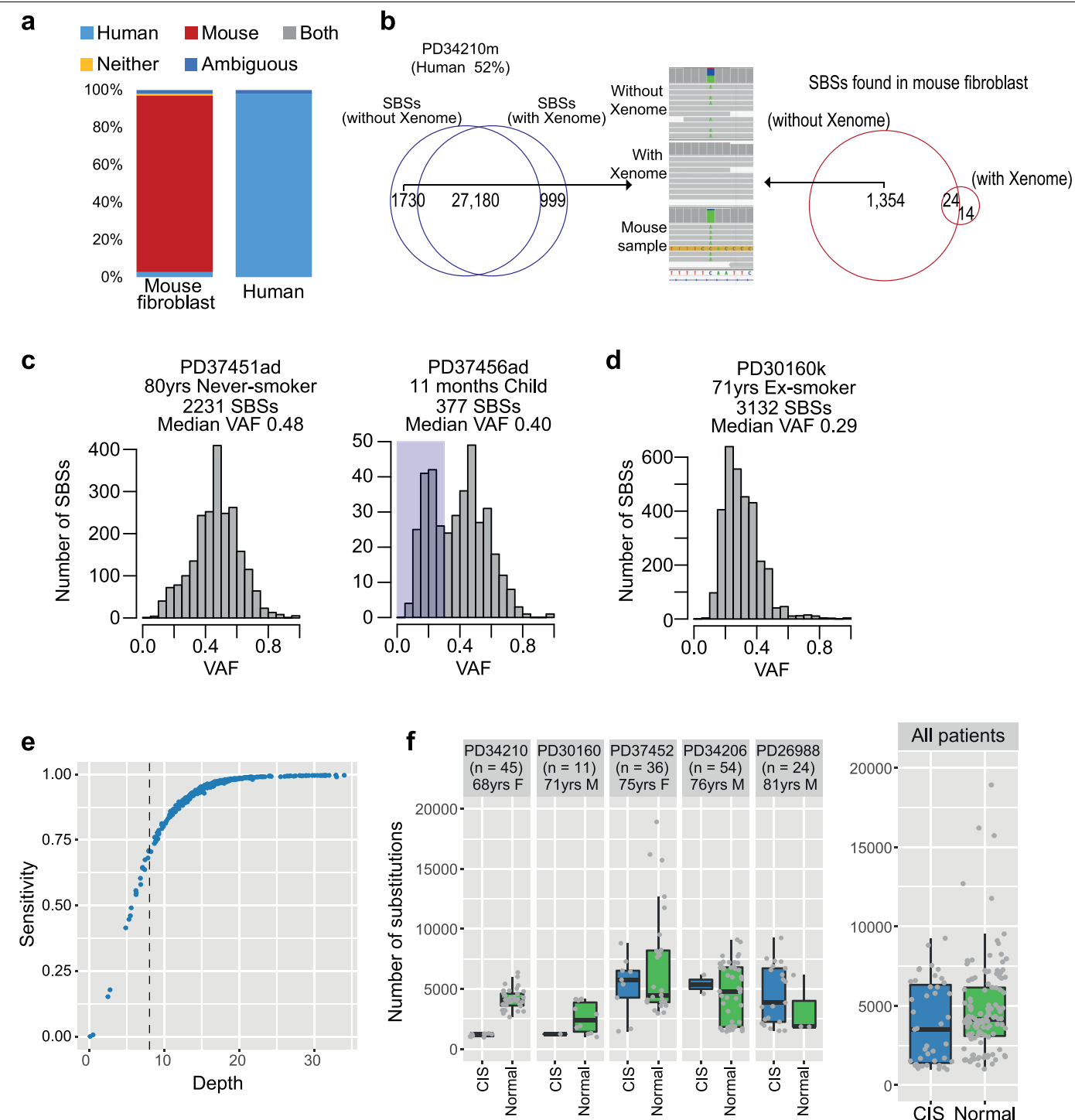
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

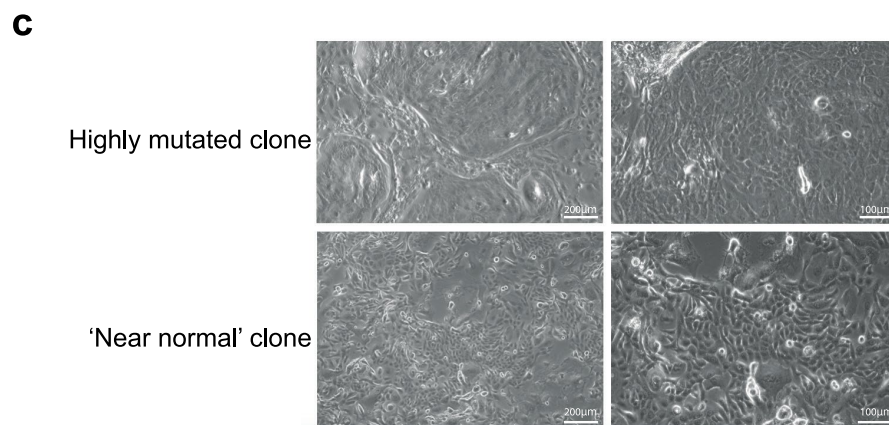
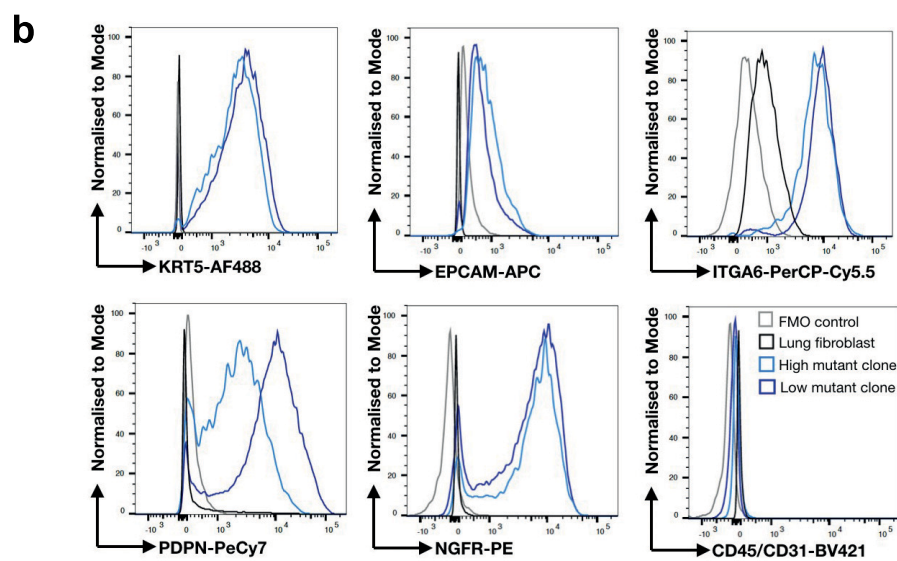
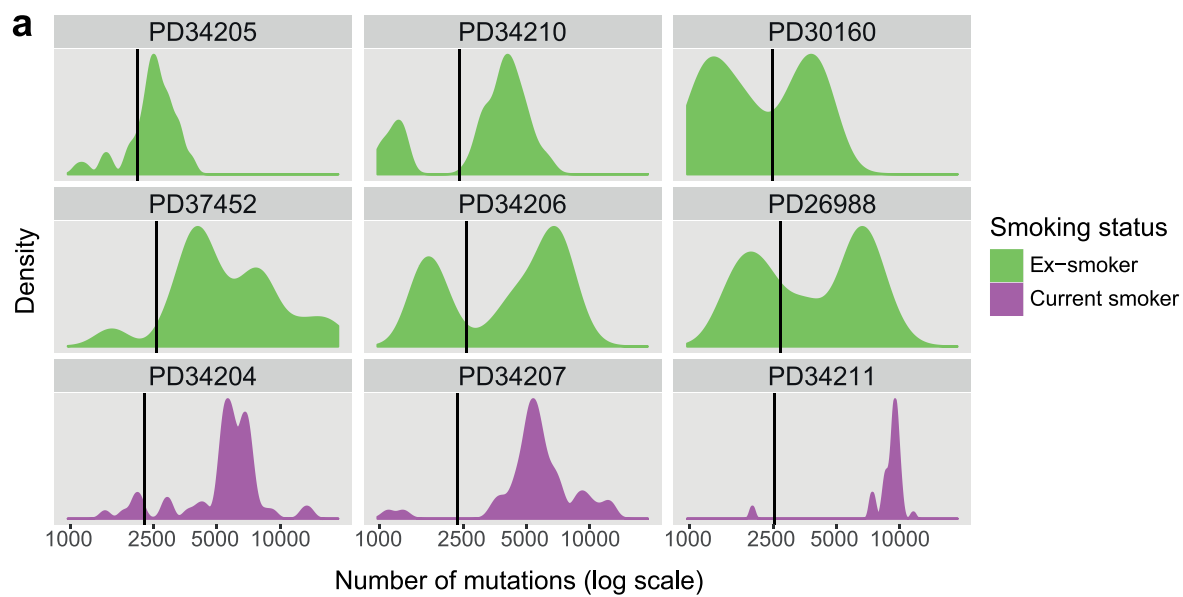
Extended Data Fig. 1 | Flow-sorting strategy of single basal bronchial epithelial cells. a, Sorting of EpCAM⁺ epithelial cells from human airway biopsies. Human haematopoietic and endothelial cells were stained with antibodies against CD45 and CD31, respectively. Within the population of cells negative for those markers, EpCAM-expressing cells were gated. Single, live (DAPI-negative) cells were flow-sorted from this population into individual wells of 96-well plates. **b,** Quantitative PCR (qPCR) analysis of cultures of clonally derived airway epithelial cells. Airway basal cells express integrin subunit $\alpha 6$ (*ITGA6*), keratin 5 (*KRT5*), cadherin 1 (*CDH1*) and *TP63*. Expression is shown in clonally derived cell cultures ($n=13$ from 3 donors, coloured blue,

green and orange) compared to control bulk human bronchial epithelial cell cultures (HBECS) that were expanded in the same culture conditions and lung fibroblast cell cultures (lung fibs) that served as a negative control. The centre values and error bars indicate mean and s.e.m., respectively. Conditions in which no expression was detected are shown as 0. **c,** Colony-forming efficiency of CD45⁺ CD31⁺ EPCAM⁺ cells after single-cell sorting from endobronchial biopsy samples ($n=16$). For one ex-smoker, EpCAM was not used to select cells and only CD45⁺ CD31⁺ cells were sorted; as expected, this was the patient with the lowest colony-forming efficiency.



Extended Data Fig. 2 | Quality assurance of mutation calls. **a**, Stacked bar chart showing the proportion of reads attributed to the human genome, mouse genome, both, neither, or with ambiguous mapping for the pure mouse fibroblast feeder line (left) or a pure human sample (right), assessed with the Xenome pipeline. **b**, Clean-up of mutation calls using the Xenome pipeline for one of the samples that was more heavily contaminated by the mouse feeder layer. The Venn diagram on the left shows the overlap in mutation calls before and after removing non-human reads by Xenome. **c**, Histograms of VAFs for two representative colonies in the sample set. The plot on the left shows a tight distribution around 50%, as expected for a colony derived from a single cell without contamination. The plot on the right shows a bimodal distribution with one peak at 50% (mutations present in the original basal cell) and a second peak at around 25% (probably representing mutations that were acquired in vitro

during colony expansion). These second peaks at less than 50% are more evident in colonies from children, owing to the low number of mutations in the original basal cell. **d**, Histogram of VAFs for a colony seeded by more than one basal cell, leading to a peak at much less than 50%. **e**, Estimated sensitivity of mutation calling according to sequencing depth. Heterozygous germline polymorphisms were identified in each subject; for each colony sequenced, we calculated the fraction of these polymorphisms that was recalled by our algorithms. **f**, Comparison of mutational burden in normal bronchial epithelial cells that neighbour a carcinoma in situ (CIS) versus cells distant from the CIS in five patients. The box-and-whisker plots show the distribution of mutational burden per colony within each subject, with the boxes indicating median and interquartile range and the whiskers denoting the range. The overlaid points are the observed mutational burden of individual colonies.



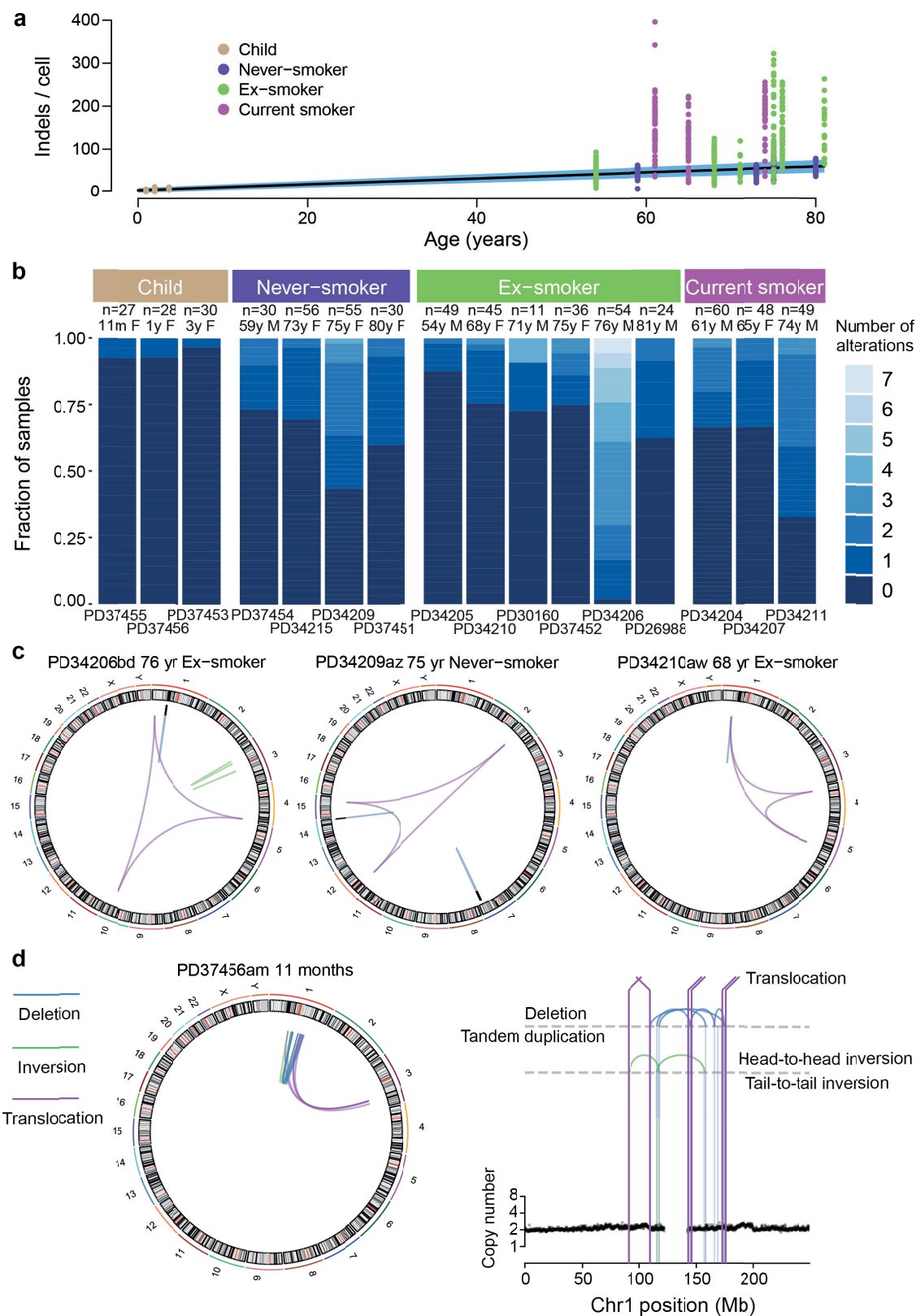
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Colonies with a near-normal mutational burden.

a, Density distribution of mutational burden in cells from ex-smokers (green) and current smokers (purple). The black vertical line shows the threshold for near-normal mutational burden derived for each patient. The x axis is on a logarithmic scale. Note the frequently bimodal distribution of mutational burden, especially in the ex-smokers, with the modes separated at the threshold for near-normal mutational burden. **b**, Flow cytometric analysis of clones for expression of KRT5, EpCAM, ITGA6, podoplanin (PDPN), NGFR and CD45 or CD31. Lung fibroblasts are included as a comparison. Fluorescence minus one (FMO) is shown. Plots for one clone with a near-normal mutational

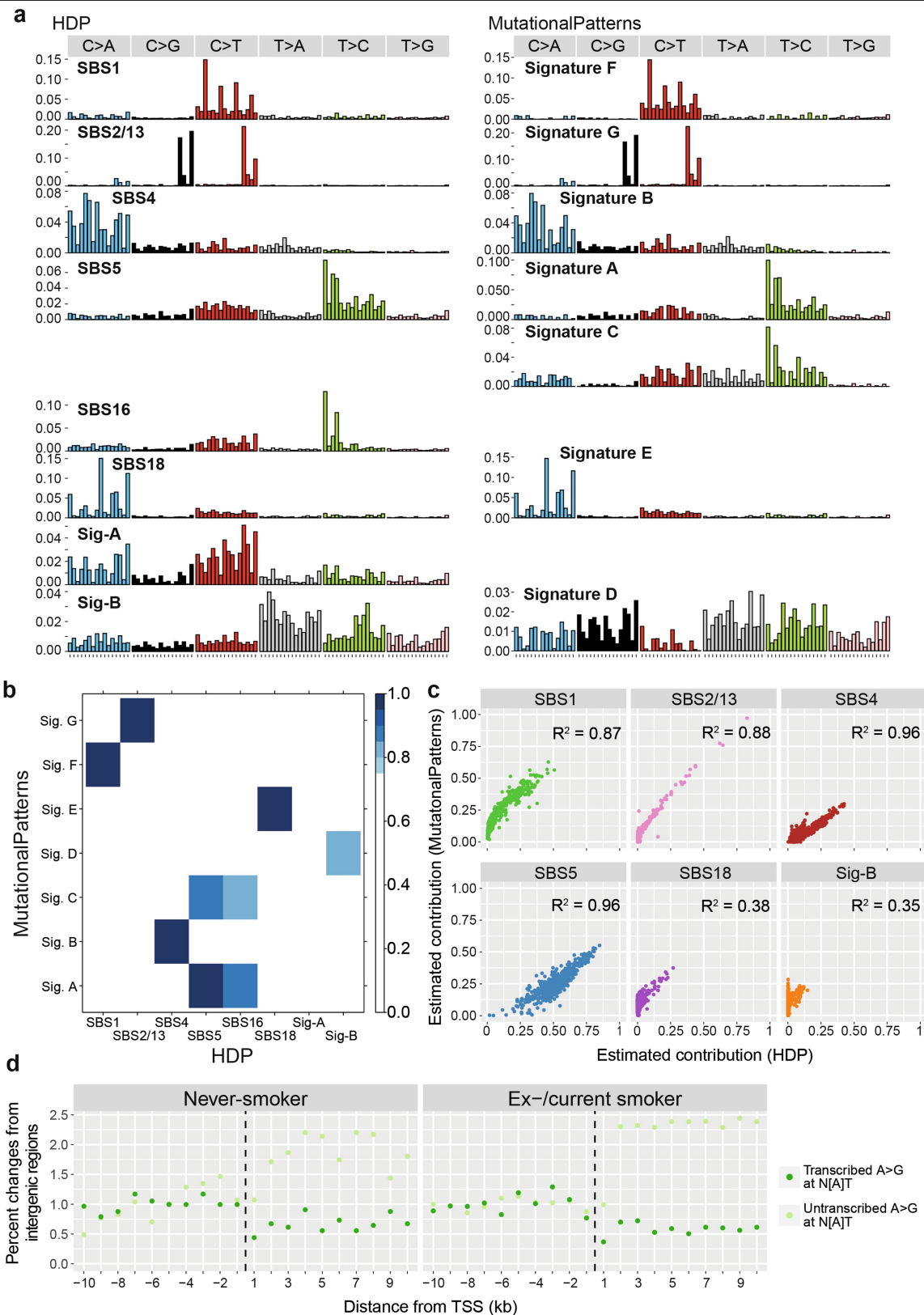
burden (low-mutant clone) and one with an increased burden (high-mutant clone) are shown, and are representative of five clones from one patient.

c, Bright-field images of expanded clones at passage 3, showing cobblestone epithelial morphology. Images are representative of five clones from one patient. A clone with an increased mutational burden is shown at the top, and a clone from an ex-smoker with a near-normal mutational burden is shown at the bottom. For the left images, the magnification is $\times 10$ and the scale bar is $200\text{ }\mu\text{m}$; for the right images, the magnification is $\times 20$ and the scale bar is $100\text{ }\mu\text{m}$.



Extended Data Fig. 4 | Indels, copy-number changes and structural variants in normal bronchial epithelial cells. a, Relationship of burden of indels per cell with age. The points represent individual colonies ($n = 632$) and are coloured by smoking status. The black line represents the fitted effect of age on indel burden, which was estimated from LME models after correction for smoking status and within-patient correlation structure. The blue shaded area represents the 95% CI for the fitted line. **b**, Stacked bar plot showing the distribution of colonies with 0–7 copy-number changes and structural variants

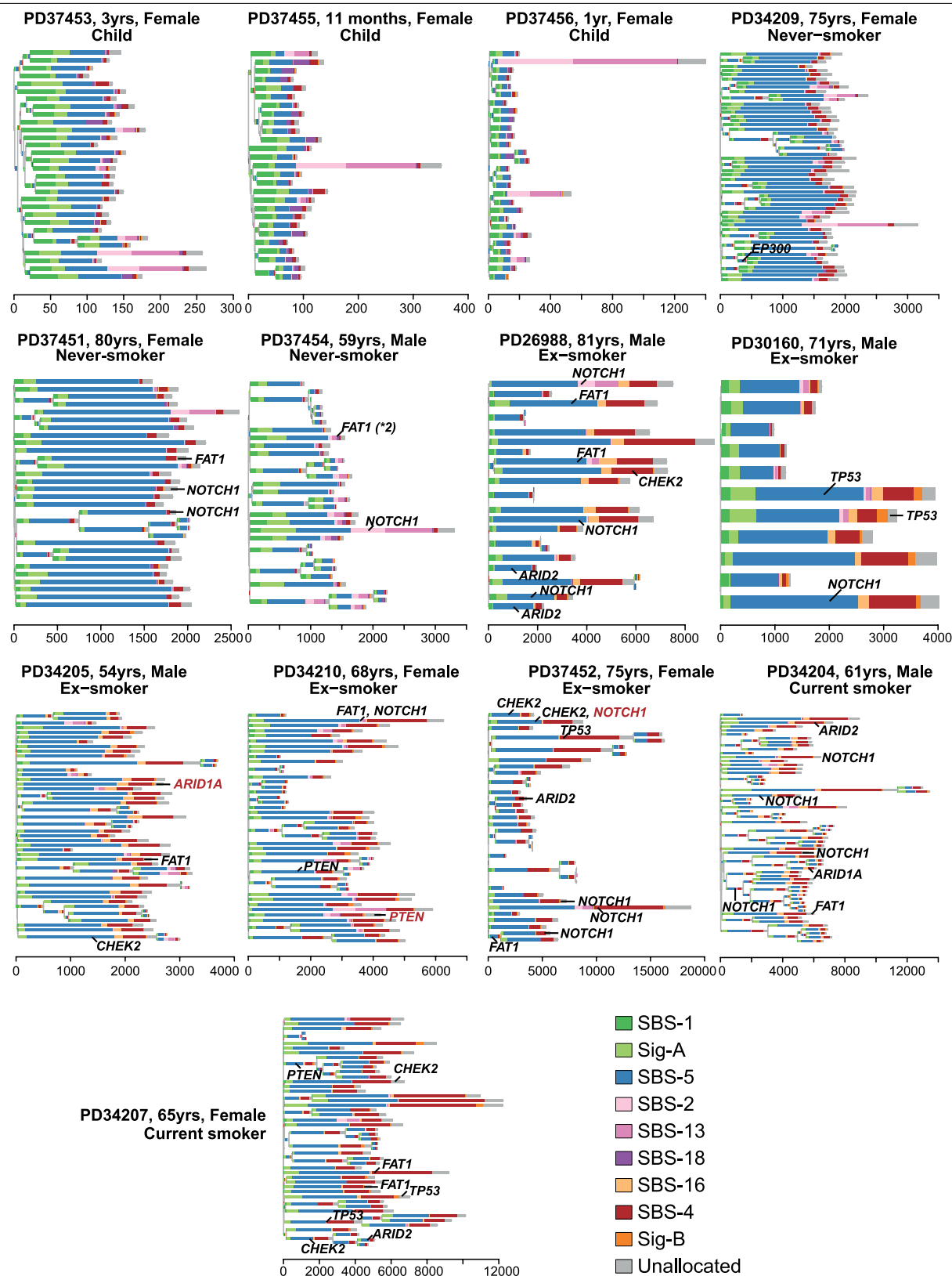
across the 16 subjects. **c**, Three examples of chromoplexy in normal bronchial cells. Structural variants are shown as coloured arcs that join two positions in the genome around the circumference. The instances of chromoplexy all consist of three translocations (purple). **d**, An example of chromothripsis in a cell from an 11-month old child. The plot on the right shows the copy number of genomic windows in the relevant region of chromosome 1 (black points); the lines and arcs denote the positions of observed structural variants.



Extended Data Fig. 5 | See next page for caption.

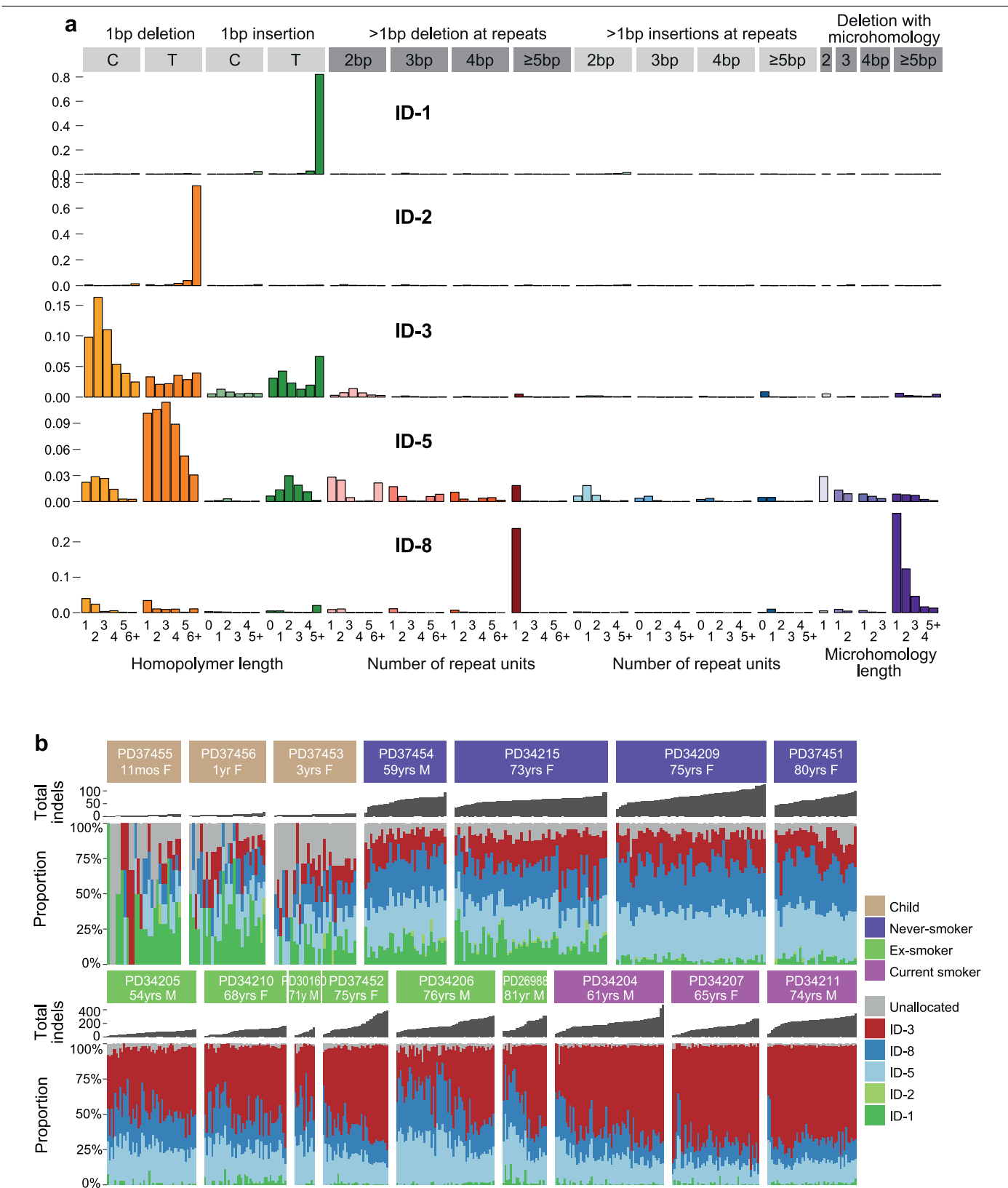
Extended Data Fig. 5 | Comparison of mutational signatures that were extracted using two algorithms. a, Trinucleotide contexts for the signatures extracted by the hierarchical Dirichlet process (HDP) (left) and MutationalPatterns non-negative matrix factorization (right). The six substitution types are shown across the top of each signature. Within each signature, the trinucleotide context is shown as four sets of four bars, grouped by whether an A, C, G or T respectively is 5' to the mutated base, and within each group of four by whether A, C, G or T is 3' to the mutated base (the order of bars is the same as that shown in Fig. 2b). Where signatures show high cosine similarity scores between algorithms, they are lined up horizontally. We note that Signature C in MutationalPatterns does not have a match in the signatures extracted by the HDP algorithm, but appears very similar to Signature A in

MutationalPatterns (or SBS-5 from the HDP). This means that it probably represents over-splitting of the signatures. **b,** Heat map showing the cosine similarities of signatures extracted by MutationalPatterns with those extracted by the HDP. Only cosine-similarity scores that are greater than 0.75 are coloured. **c,** Scatter plots showing the fraction of mutations in each colony ($n = 632$) assigned to each signature by the HDP algorithm (x axis) versus the MutationalPatterns algorithm (y axis). The correlation values quoted are Pearson's correlation coefficients (R^2). **d,** Transcriptional strand bias of A>G mutations in an N[A]T context before and after TSSs. Note the absence of transcriptional strand bias in intergenic regions but evidence for both transcription-coupled damage and repair after the TSS, applying similarly in both never-smokers and ex- or current smokers.



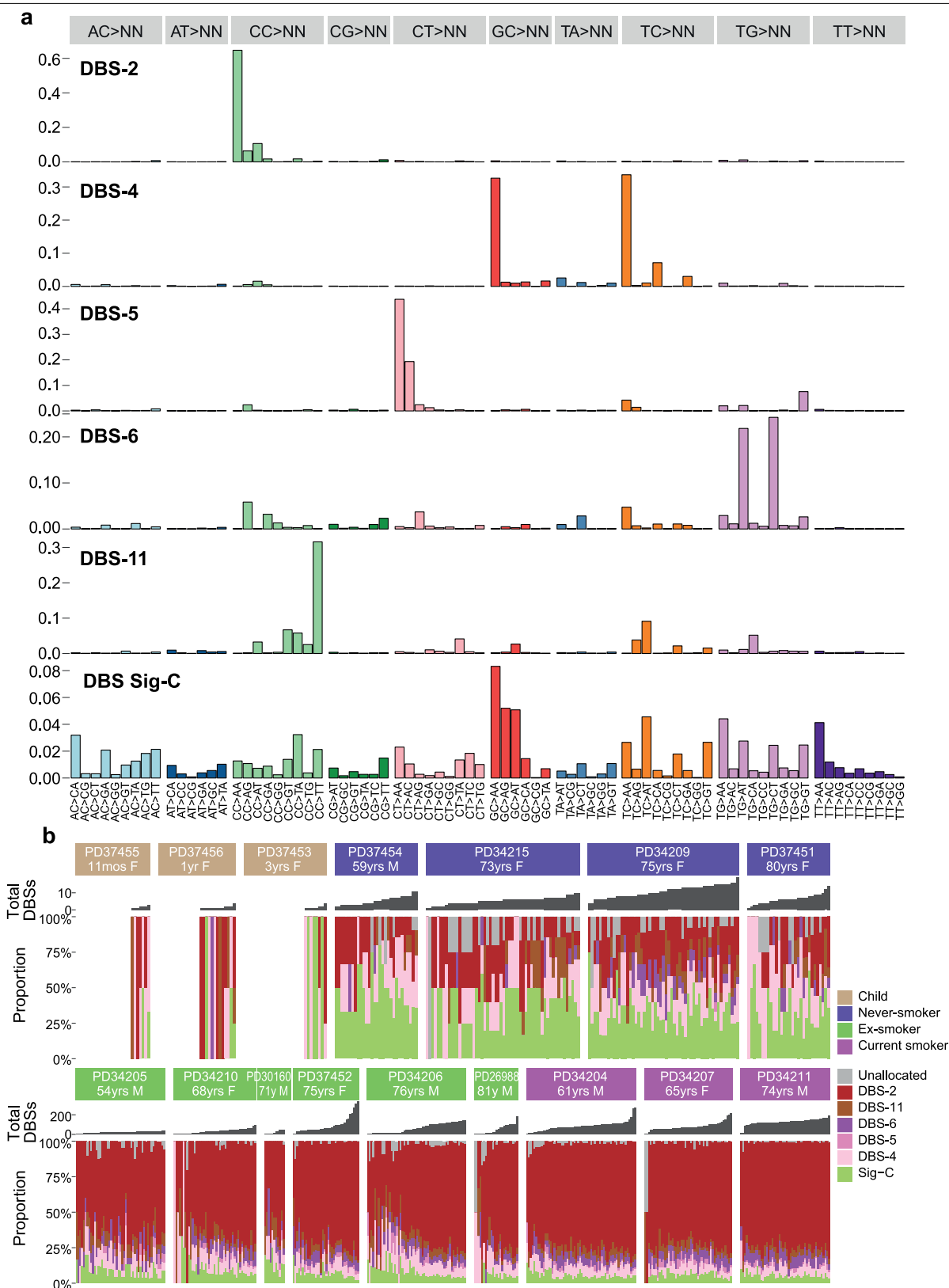
Extended Data Fig. 6 | Phylogenetic trees of 13 subjects. Phylogenetic trees showing clonal relationships among normal bronchial cells in the 13 subjects not shown in Fig. 3a. Branch lengths are proportional to the number of

mutations (x axis) specific to that clone or subclone. Each branch is coloured by the proportion of mutations on that branch that are attributed to the various SBS signatures.



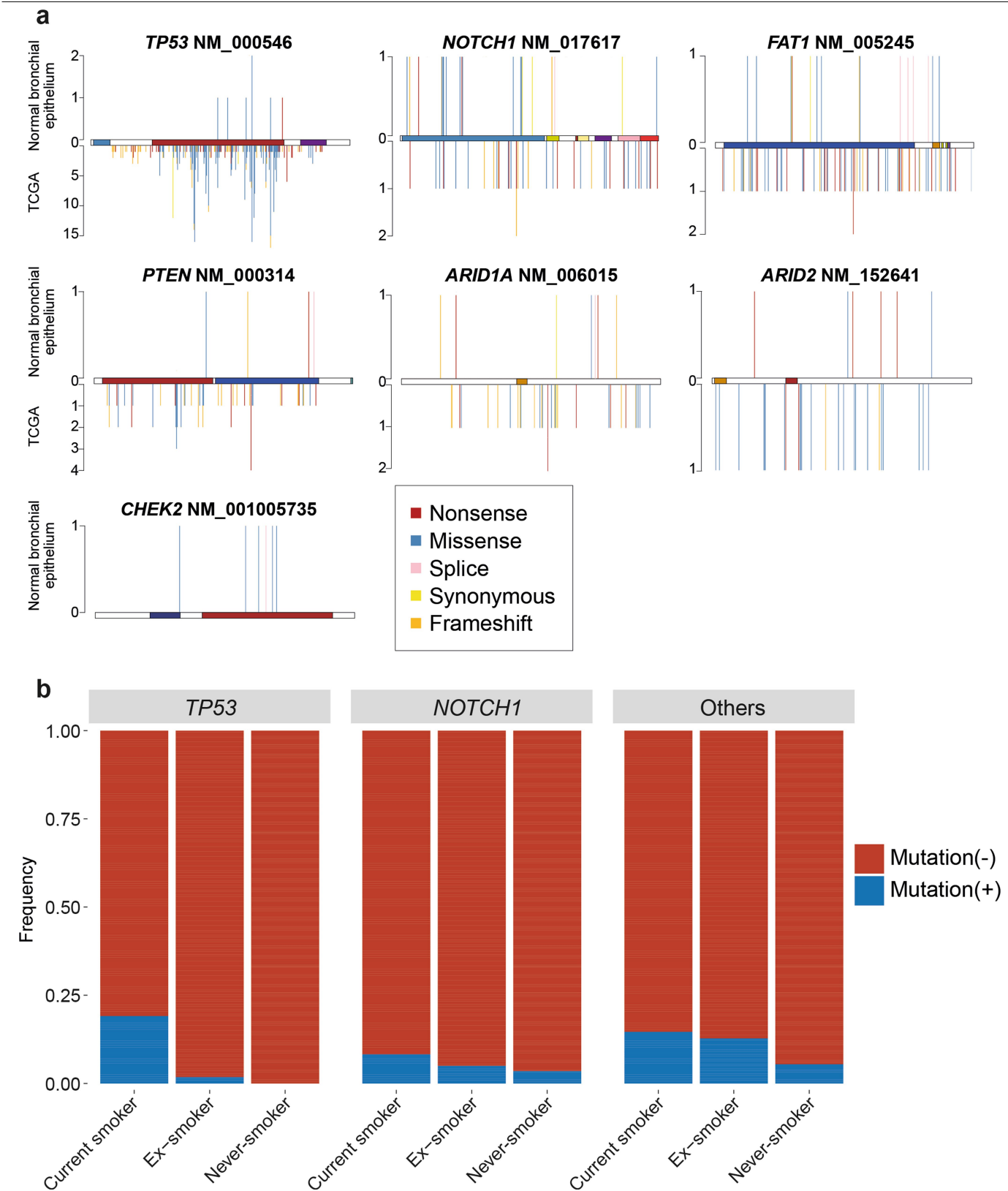
Extended Data Fig. 7 | Indel signatures in the sample set. a. Five indel signatures (ID-1, ID-2, ID-3, ID-5 and ID-8) were extracted by the HDP. The contributions of different types of indels to each signature are shown, grouped by whether variants are deletions or insertions; the size of the event; whether they occur at repeat units; and the sequence content of the indel. **b.** Stacked bar

plot showing the proportional contribution of mutational signatures to indels across the 632 colonies derived from normal bronchial cells, extracted using the HDP. Within each patient, colonies are sorted from left to right by increasing indel burden (bar chart in dark grey above coloured signature-attribution stacks).



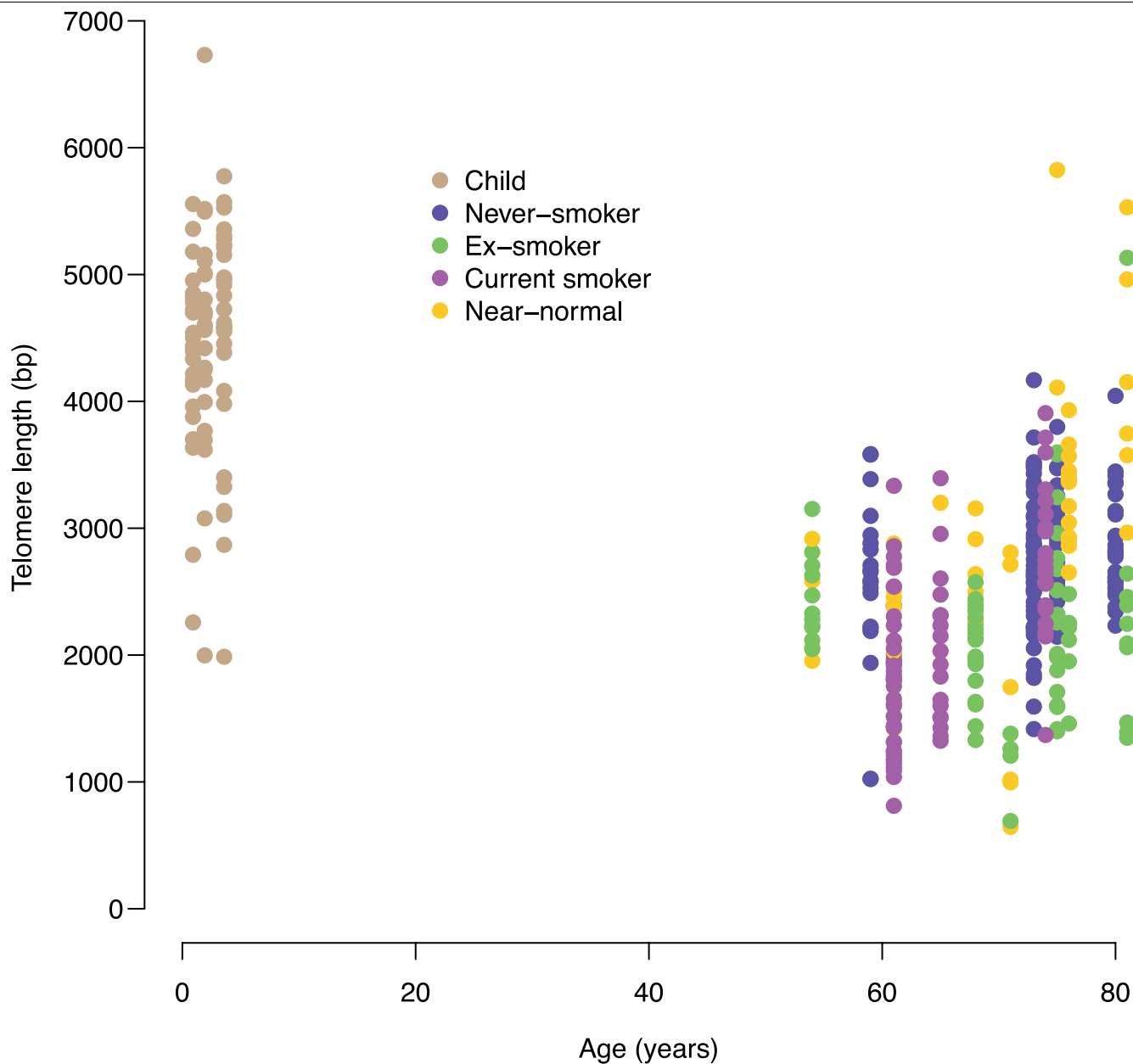
Extended Data Fig. 8 | DBS signatures in the sample set. a, Six DBS signatures were extracted by the HDP. The contributions of different types of double-base substitution to each signature are shown, grouped by the sequence that is mutated and by what it is mutated to. Five of the signatures have been observed in cancer genomes²⁴, and one (DBS Sig-C) is a novel signature that was extracted here. **b,** Stacked bar plot showing the proportional contribution of mutational

signatures to double-base substitutions across the 632 normal bronchial cells, extracted using the HDP. Note that some of the colonies in children have no double-base substitutions. Within each patient, colonies are sorted from left to right by increasing burden of double-base substitutions (bar chart in dark grey above coloured signature-attribution stacks).



Extended Data Fig. 9 | Driver mutations in normal bronchial epithelium.
a, Stick plots showing distribution of mutations in *TP53*, *NOTCH1* and other genes that were significantly mutated in our sample set. Mutations are coloured by type. The gene structure is shown horizontally in the centre of each plot, with domains as coloured bars. Above the gene are mutations in this

sample set, and below the gene are mutations found in squamous cell carcinomas from the TCGA sample set. **b**, Fraction of cells with driver mutations in *TP53* (left), *NOTCH1* (middle) or all other significant cancer genes (right), split by smoking status.



Extended Data Fig. 10 | Relationship of telomere length with age. Scatter plot of estimated telomere lengths (y axis) against the age of the subject (x axis). Individual points represent colonies ($n = 398$ colonies in which less

than 10% of the DNA was derived from the mouse feeder layer). Cells with a near-normal mutational burden are coloured gold.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Image processing from sequencing data using standard Illumina X10 pipeline

Data analysis

Alignment and variant calling performed using Sanger Institute's custom pipeline. Single-nucleotide substitutions were called using the CaVEMan (cancer variants through expectation maximization) algorithm (<https://github.com/cancerit/CaVEMan>). Small insertions and deletions were called using the Pindel algorithm (<https://github.com/genome/pindel>). Rearrangements were called using the BRASS (breakpoint via assembly) algorithm (<https://github.com/cancerit/BRASS>).

List of programs and softwares:

R – version 3.5.1

BWA-MEM - version 0.7.17-r1188 (<https://sourceforge.net/projects/bio-bwa/>)

CaVEMan - version 1.11.2

Pindel - version 2.2.5

Brass - version 6.1.2

ASCAT NGS - version 4.1.2

Xenome (<https://github.com/data61/gossamer/blob/master/docs/xenome.md>)

deepSNV - version 1.28.0 (<https://bioconductor.org/packages/release/bioc/html/deepSNV.html>)

ANNOVAR (<http://wannovar.wglab.org/>)

IGV (<http://software.broadinstitute.org/software/igv/>)

JBrowse (<https://jbrowse.org/>)

cgpVAF (<https://github.com/cancerit/vafCorrect>)RPhylip - version 0.1.23 (<http://www.phytools.org/Rphylip/>)hdp - version 0.1.5 (<https://github.com/nicolaroberts/hdp>)MutationalPatterns - version 1.8.0 (<https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html>)dNdScv - version 0.0.1 (<https://github.com/im3sanger/dndscv>)Telomerecat - version 3.1.2 (<https://github.com/jhrf/telomerecat>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that support the findings of this study have been deposited in the European Genome-Phenome Archive (<https://www.ebi.ac.uk/ega/home>) under accession number EGAD00001005193. Somatic mutation calls, including single base substitutions, indels and structural variants, from all 632 samples have been deposited on Mendeley Data with the identifier: <http://dx.doi.org/10.17632/b53h2kwpyy.2>.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen to give good representation of inter-patient and intra-patient variability in mutation burden.
Data exclusions	Oligoclonal colonies or with low mean coverage (<8x) (Extended Figure 2e) were excluded due to the inaccuracy of mutation catalogues. One outlying cell in an ex-smoker with >10,000 mutations is excluded from the plot of Figure 4 to improve visualisation.
Replication	No experimental replication has yet been attempted.
Randomization	Not applicable - this is a descriptive study, not an intervention study.
Blinding	Not applicable - all dependent variables were computationally generated (mutation counts, signatures etc) and statistical analyses were prespecified.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used Antibodies used in this study were described in OnlineMethods.

Validation

Antibodies were validated by the manufacturer.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

3T3-J2 feeder cells were kindly provided by Prof. Fiona Watt (King's College London)

Authentication

The feeder cells had whole genome sequencing performed, confirming their murine origin and clonal derivation.

Mycoplasma contamination

They were tested negative for Mycoplasma by PCR test (PMCID: PMC202165)

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used in this study.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

We analysed single cell-derived colonies from bronchial epithelium of 16 subjects, including 3 children, 4 never-smokers, 6 ex-smokers and 3 current smokers. Clinical characteristics of the cohort are described in Supplementary Table 1. Of the ex-smokers, 2 had had a previous cancer treated with curative intent, and 5 had a carcinoma in situ or invasive squamous cell carcinoma. The children in the cohort had bronchoscopy for investigation or follow-up of congenital anomalies.

Recruitment

Recruited through University College Hospitals, London, UK. Our cohort does potentially suffer from recruitment bias, since samples could only ethically be obtained from individuals undergoing a clinically indicated bronchoscopy.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

The epithelium was dissected away from the underlying stroma and fetal bovine serum (FBS) was added to a final concentration of 10%. Both the epithelium and stroma were combined and digested in 0.1% trypsin/EDTA at 37 °C for 30 minutes. The solution was neutralised with FBS to a final concentration of 10% and added to the neutralised dispase solution¹. Cells were passed through a 100 µm cell strainer and stained in sorting buffer (1x PBS, 1% FBS, 25 mM HEPES and 1 mM EDTA) with anti-CD45-PE (BD Pharmingen 555483, 1:200), anti-CD31-PE (BD Pharmingen 555446, 1:200), anti-EPCAM-APC (Biolegend 324208, 1:50) antibodies and DAPI (1 µg/ml).

Instrument

BD FACSAria Fusion

Software

Supplementary Methods

Cell population abundance

Supplementary Methods

Gating strategy

Supplementary Methods

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Discriminating α -synuclein strains in Parkinson's disease and multiple system atrophy

<https://doi.org/10.1038/s41586-020-1984-7>

Received: 14 November 2018

Accepted: 10 January 2020

Published online: 5 February 2020

Mohammad Shahnawaz¹, Abhisek Mukherjee¹, Sandra Pritzkow^{1,6}, Nicolas Mendez^{1,6}, Prakruti Rabadia¹, Xiangnan Liu², Bo Hu², Ann Schmeichel³, Wolfgang Singer³, Gang Wu⁴, Ah-Lim Tsai⁴, Hamid Shirani⁵, K. Peter R. Nilsson⁵, Phillip A. Low³ & Claudio Soto^{1*}

Synucleinopathies are neurodegenerative diseases that are associated with the misfolding and aggregation of α -synuclein, including Parkinson's disease, dementia with Lewy bodies and multiple system atrophy¹. Clinically, it is challenging to differentiate Parkinson's disease and multiple system atrophy, especially at the early stages of disease². Aggregates of α -synuclein in distinct synucleinopathies have been proposed to represent different conformational strains of α -synuclein that can self-propagate and spread from cell to cell^{3–6}. Protein misfolding cyclic amplification (PMCA) is a technique that has previously been used to detect α -synuclein aggregates in samples of cerebrospinal fluid with high sensitivity and specificity^{7,8}. Here we show that the α -synuclein-PMCA assay can discriminate between samples of cerebrospinal fluid from patients diagnosed with Parkinson's disease and samples from patients with multiple system atrophy, with an overall sensitivity of 95.4%. We used a combination of biochemical, biophysical and biological methods to analyse the product of α -synuclein-PMCA, and found that the characteristics of the α -synuclein aggregates in the cerebrospinal fluid could be used to readily distinguish between Parkinson's disease and multiple system atrophy. We also found that the properties of aggregates that were amplified from the cerebrospinal fluid were similar to those of aggregates that were amplified from the brain. These findings suggest that α -synuclein aggregates that are associated with Parkinson's disease and multiple system atrophy correspond to different conformational strains of α -synuclein, which can be amplified and detected by α -synuclein-PMCA. Our results may help to improve our understanding of the mechanism of α -synuclein misfolding and the structures of the aggregates that are implicated in different synucleinopathies, and may also enable the development of a biochemical assay to discriminate between Parkinson's disease and multiple system atrophy.

The misfolding and aggregation of α -synuclein (α -syn) involves a mechanism of seeding and nucleation, in which initial seeds of α -syn recruit other soluble monomers that assemble to form aggregates^{9,10}. Aggregates of α -syn circulate in biological fluids such as the cerebrospinal fluid (CSF) and blood^{11,12}. The process of protein misfolding and aggregation appears to begin years or decades before the onset of clinical signs, and thus detection of α -syn aggregates in easily accessible biological fluids may enable the biochemical diagnosis of synucleinopathies. In previous studies, the PMCA technology has been adapted to enable highly sensitive and specific detection of α -syn aggregates that are produced *in vitro*^{6,13,14} or derived from the biological fluids of

patients with synucleinopathies^{7,8}. The α -syn-PMCA assay (also referred to as α -syn-RT-QuIC^{15,16}) uses the seeding–nucleation mechanism to cyclically amplify the process of protein misfolding, enabling the efficient amplification of small quantities of α -syn oligomers and thereby facilitating their detection.

In the α -syn-PMCA assay, the kinetics of aggregation of α -syn are monitored by the fluorescence signal of thioflavin T (ThT)—a dye that is specific to amyloid fibrils¹⁷. Previous studies have noted that the maximum fluorescence signal of the α -syn-PMCA product from reactions that were initiated with CSF from patients with multiple system atrophy (MSA) was smaller than the corresponding fluorescence signal

¹Mitchell Center for Alzheimer's Disease and Related Brain Disorders, Department of Neurology, University of Texas McGovern Medical School at Houston, Houston, TX, USA. ²Department of Microbiology and Molecular Genetics, University of Texas McGovern Medical School at Houston, Houston, TX, USA. ³Department of Neurology, Mayo Clinic, Rochester, MN, USA. ⁴Division of Hematology, Department of Internal Medicine, University of Texas McGovern Medical School at Houston, Houston, TX, USA. ⁵Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden. ⁶These authors contributed equally: Sandra Pritzkow, Nicolas Mendez. *e-mail: Claudio.Soto@uth.tmc.edu

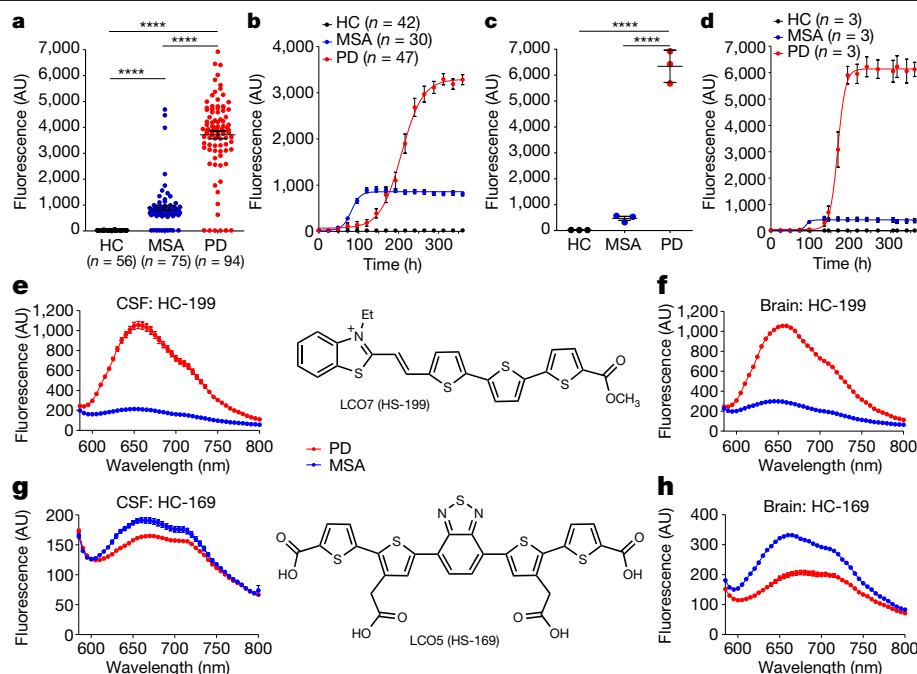


Fig. 1 | Differential interaction of amyloid-binding dyes with α -syn aggregates derived from patients with PD or patients with MSA.

a, b, Samples of CSF (40 μ l) from patients with PD (PD), patients with MSA or healthy control individuals (HC) were subjected to α -syn-PMCA and the extent of aggregation was monitored by ThT fluorescence. **a**, Maximum fluorescence values (measured at plateau of aggregation) for PD ($n = 94$; red), MSA ($n = 75$; blue) and healthy controls ($n = 56$; black). Each dot represents an individual biological sample measured in duplicate and data are mean \pm s.e.m. **b**, Representative aggregation curves of α -syn in the presence of CSF from patients with PD ($n = 47$), patients with MSA ($n = 30$) and healthy controls ($n = 42$). Data are mean \pm s.e.m. of all patients analysed in each group. **c, d**, Frozen brain samples from patients with pathologically confirmed PD or MSA, or from healthy controls, were homogenized at 10% w/v. A 0.001% dilution of brain homogenate was used for the α -syn-PMCA reaction.

c, Maximum fluorescence values for PD ($n = 3$), MSA ($n = 3$) and healthy controls ($n = 3$). Each dot represents an individual biological sample measured in duplicate and data are mean \pm s.e.m. of three patients in each group.

**** $P < 0.0001$ by one-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test (**a, c, d**). Aggregation profiles of α -syn in the presence of samples from the brain of patients with PD ($n = 3$), patients with MSA ($n = 3$) and healthy controls ($n = 3$). Data are mean \pm s.e.m. of three patients in each group. **e–h**, Differential binding of two amyloid-conformation-specific dyes (HS-199 and HS-169) to α -syn aggregates obtained after two rounds of α -syn-PMCA in samples from the CSF (**e, g**; $n = 43$) or the brain (**f, h**; $n = 3$) of different patients with PD or MSA. Excitation was at 540 nm and the emission spectrum was recorded between 580 and 800 nm. The chemical structures of HS-199 and HS-169 are also shown. Each experiment was performed in duplicate and data are mean \pm s.e.m. (for many points the error bars are smaller than the symbols).

for CSF from patients with Parkinson's disease (PD) or dementia with Lewy bodies⁷. To further investigate the possibility that PD and MSA can be differentiated by α -syn-PMCA, we performed a study using 94 samples of CSF from patients with PD, 75 from patients with MSA and 56 from control individuals with other neurological diseases (Methods; see Extended Data Table 1 for patient demographics). The maximum ThT fluorescence after α -syn-PMCA was significantly greater in samples from patients with PD than in samples from patients with MSA (Fig. 1a). Products of α -syn-PMCA that were derived from samples from patients with MSA had a maximum fluorescence of less than 1,800 units, whereas for PD this value ranged between 2,000 and 8,000 units. Control samples did not show any fluorescence over the background levels (Fig. 1a). The kinetics of aggregation for all samples in this study are shown in Extended Data Fig. 1. Of the 75 samples from patients with MSA, 4 had an aggregation profile that was compatible with the PD strain and, conversely, 3 of the 94 samples from patients with PD had a profile typical of MSA. From this cohort of samples the overall sensitivity for diagnosis of PD and MSA, as compared to controls calculated by receiving operating curves, was 93.6% and 84.6%, respectively. In both cases, specificity was 100%. Comparing differential diagnosis of PD and MSA, we estimated that of the 88 samples from patients with clinically diagnosed PD that showed α -syn seeds by α -syn-PMCA, 85 were correctly identified as PD in our assay (that is, a sensitivity of 96.6%). Of the 65 samples from patients with MSA that were shown by α -syn-PMCA to contain α -syn aggregates, 61 had the typical signature of MSA (maximum fluorescence of less than 1,800), indicating a sensitivity

of 93.8%. Combining all samples, we correctly distinguished PD from MSA in 146 of the 153 samples analysed—an overall sensitivity of 95.4%.

The above data were obtained from different cohorts of patients and across several separate experiments. To illustrate the typical profile of α -syn-PMCA aggregation for samples of PD and MSA, we took the largest individual cohort of samples analysed in Fig. 1a and plotted data from samples that were identified as PD ($n = 47$) and MSA ($n = 30$) (Fig. 1b). The maximum fluorescence and the kinetics of aggregation were consistently different for PD and MSA, with samples from patients with MSA aggregating faster but reaching a lower fluorescence plateau than those from patients with PD (Fig. 1b). To determine whether the aggregates present in the CSF are representative of those found in the brain, we also amplified brain samples from three different patients with PD or MSA. To reduce the chance of other brain components interfering in the reaction, we started the PMCA assay with a 10^{-4} dilution of brain homogenate. Under these conditions, we found that amplified brain-derived α -syn aggregates showed the typical signature of PD or MSA, both in terms of the maximum ThT fluorescence (Fig. 1c) and the kinetics of aggregation (Fig. 1d). These results suggest that the aggregates present in the CSF of patients reflect the aggregates present in the brain.

Notably, the qualitative differences in ThT fluorescence were maintained when the α -syn aggregates that were amplified from samples of CSF from patients with PD or patients with MSA were replicated serially at the expense of monomeric α -syn (Extended Data Fig. 2). For these studies, an aliquot of the final product of the first α -syn-PMCA reaction (starting from CSF samples) was diluted 100-fold into fresh

α -syn monomers, and a new α -syn-PMCA assay was performed. This was repeated several times, and the product maintained the high-fluorescence signal for PD and low-fluorescence signal for MSA (Extended Data Fig. 2). To further study the properties of the aggregates that were amplified from patients with PD or with MSA, we selected samples from 43 patients with PD and 43 patients with MSA (see Extended Data Table 2 for the demographic characteristics of these patients). The selection of the 43 samples for each disease was done by eliminating samples that did not aggregate (false negatives) and including those that had the typical signatures of PD or MSA, as indicated above (Fig. 1b, Extended Data Fig. 1). The majority of the characterization studies were done with samples from the second cycle of amplification; this was necessary to generate sufficient material and also to reduce any interference from the CSF, which is important for some of the techniques used (for example, circular dichroism and Fourier-transform infrared (FTIR) spectroscopy).

First, we wanted to verify that the differences in ThT fluorescence did not simply reflect different amounts of aggregates at the end of the reaction. To investigate this further, we performed sedimentation assays to separate the pools of soluble and aggregated α -syn. We measured the amount of protein pelleting after centrifugation at 20,000g for 30 minutes, using silver staining after SDS-PAGE (Extended Data Fig. 3a) and dot blot analysis (Extended Data Fig. 3b). We also measured the amount of protein remaining in the supernatant, using the bicinchoninic acid assay (Extended Data Fig. 3c). The results clearly showed that the amount of aggregates produced at the end of the α -syn-PMCA assay was the same in both the PD and the MSA samples. Our interpretation of these results is that either the accessibility or the mode of interaction of ThT with aggregates differs between aggregates derived from patients with PD and those derived from patients with MSA, and that this probably reflects structural differences in the aggregates.

To study the differences between aggregates associated with PD and aggregates associated with MSA in more detail, we first used a panel of thiophene-based ligands that have previously been shown to interact with amyloid aggregates and produce a different spectrum depending on the structural characteristics of the aggregates^{18,19}. The conjugated thiophene backbone is flexible and thus the binding and fluorescence emission of the molecules depends on the conformational properties of the aggregates, providing a specific spectral fingerprint of different aggregates^{18,19}. These compounds have previously been shown to discriminate between different conformational strains of prions, amyloid β and tau proteins^{20,21}. We analysed a set of seven different thiophene-based ligands and found that some of them showed substantially different capacities to interact with α -syn aggregates derived from PD samples compared to those derived from MSA samples (Fig. 1e–h). HS-199 showed a very specific binding affinity and high emission of fluorescence for PD aggregates, whereas the fluorescence of this dye in the presence of MSA aggregates was very low (Fig. 1e). Similar results were obtained when analysing samples derived from brain extracts (Fig. 1f), further supporting the conclusion that aggregates amplified from the CSF and the brain are equivalent. Conversely, the HS-169 dye appeared to bind preferentially to MSA aggregates over PD aggregates, again in samples amplified from both the CSF (Fig. 1g) and the brain (Fig. 1h).

To analyse the biochemical differences between α -syn aggregates derived from patients with PD and from patients with MSA, we examined their resistance to proteolytic degradation and performed epitope-mapping experiments. Limited protease digestion is commonly used to distinguish prion strains²². Aggregates of α -syn derived by seeding and amplification from the CSF of patients with PD or patients with MSA differed in their extent of protease resistance and in the size of the core fragment that was resistant to degradation, as analysed by a panel of different antibodies (Fig. 2a–c; see Extended Data Fig. 4 for the study done with a larger number of samples). Aggregates of α -syn that were amplified from the CSF of patients with PD or patients with MSA were very resistant to degradation, even after treatment with a high

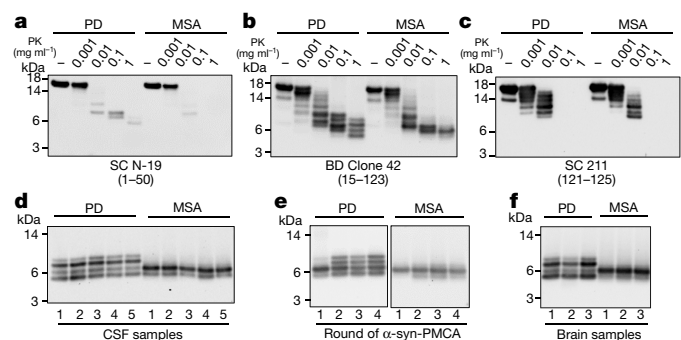


Fig. 2 | Protease resistance and epitope mapping of α -syn aggregates derived from the CSF or the brain of patients with PD or patients with MSA. **a–c**, α -syn-PMCA products starting from samples of CSF from patients with MSA or patients with PD were incubated without (–) or in the presence of increasing concentrations of proteinase K (PK; 0, 0.001, 0.01, 0.1 and 1 mg ml⁻¹) at 37 °C for 1 h. Samples were subjected to western blotting using three different antibodies against α -syn: N-19 (Santa Cruz), which recognizes the N-terminal region (residues 1–50) of α -syn (**a**); anti- α -syn clone 42 (BD Biosciences), which is raised against the middle region of α -syn (residues 15–123) (**b**); and 211 (Santa Cruz), which is reactive against the C-terminal region of α -syn (residues 121–125) (**c**). Similar results were obtained for three other patients analysed per disease (Extended Data Fig. 4). **d**, Profiles of digested fragments from five patients in each group, developed with the BD clone 42 anti- α -syn antibody. The results for all of the PD (n = 43) and MSA (n = 43) samples analysed are shown in Extended Data Fig. 5. For the experiments in **a–d**, we used the aggregates from the second round of amplification. **e**, Profile of proteinase K-resistant fragments after serial rounds of α -syn-PMCA. The first round corresponds to direct amplification from the CSF. For the second round of amplification, aggregates produced in the first round were diluted 100-fold into fresh α -syn monomer substrate and a new round of α -syn-PMCA was performed. The assay was then repeated for the third and fourth rounds using amplified α -syn aggregates (1%) from the previous round. As before, amplified aggregates were treated with proteinase K (1 mg ml⁻¹) and blots were developed with the BD clone 42 anti- α -syn antibody. **f**, Proteinase K resistance profiles of aggregates amplified from the brain of patients with neuropathologically confirmed PD (n = 3) or MSA (n = 3). Molecular weight markers (kDa) are indicated on the left of each blot.

concentration of proteinase K (1 mg ml⁻¹) for 1 hour. Under these conditions, protease-resistant fragments mostly mapped to the N-terminal (Fig. 2a) and middle (Fig. 2b) regions of the protein. Conversely, the C-terminal region of α -syn appeared to be fully degraded after incubation with more than 0.01 mg ml⁻¹ of proteinase K (Fig. 2c), which suggests that this part of the protein may not be implicated in the formation of the aggregates (consistent with previous structural studies of α -syn fibrils^{23–25}). Notably, the size and number of protease-resistant bands that were detectable by antibodies directed to the middle region of α -syn (residues 15–123) differed substantially between PD and MSA. Four bands with molecular weights ranging from 4 to 10 kDa were detected for samples from patients with PD, whereas only two bands (4 and 6 kDa) were detected for samples from patients with MSA (Fig. 2b, d). This signature was observed across all of the 43 PD and 43 MSA samples that were analysed (Fig. 2d shows 5 representative samples per disease; Extended Data Fig. 5 shows all 86 samples). The signature was maintained after serial replication *in vitro* by α -syn-PMCA (Fig. 2e, Extended Data Fig. 6), albeit with some small variability in the relative proportions of different bands between rounds of amplification. This result provides further evidence that α -syn-PMCA maintains the biochemical and structural properties of α -syn aggregates. We also analysed the pattern of proteinase K resistance of α -syn aggregates that were amplified from the brain of patients with PD or patients with MSA. The profiles of protease-resistant fragments from brain exhibited the typical signature of PD or MSA (Fig. 4d), again suggesting that the aggregates present in the CSF are equivalent to those that accumulate in the brain.

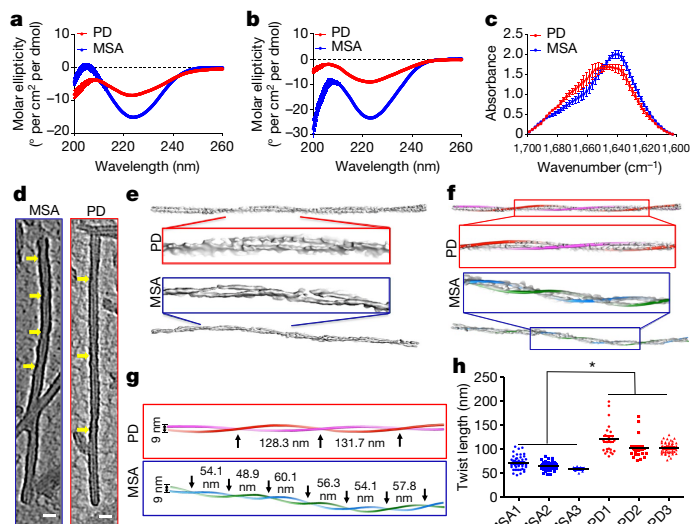


Fig. 3 | Structural differences between α -syn aggregates derived from patients with PD or patients with MSA. **a**, Circular dichroism spectra of α -syn aggregates from the CSF of patients with PD (red) or patients with MSA (blue), amplified by two rounds of α -syn-PMCA. Spectra were recorded from 35 μ M suspensions of α -syn aggregates, as described in Methods. Measurements were taken for all of the PD ($n = 43$) and MSA ($n = 43$) samples analysed and data (molar ellipticity) are mean \pm s.e.m. **b**, A similar experiment was performed for α -syn aggregates that were amplified from the brain of patients with PD ($n = 3$) or patients with MSA ($n = 3$). **c**, FTIR spectra of α -syn aggregates that were obtained after two rounds of seeding and amplification of samples of CSF from patients with PD ($n = 10$) or patients with MSA ($n = 10$). The solution of aggregated proteins (5 μ l; 5 mg ml $^{-1}$) was analysed with an FTIR-4100 spectrometer (JASCO). **d**, Cryo-ET was performed to evaluate structural differences between fibrils from patients with PD and fibrils from patients with MSA. Central slices of representative subtomograms of PD-associated fibrils and MSA-associated fibrils are shown. The negative-stained fibrils were imaged with a 300-kV electron microscope (Methods). Yellow arrows indicate twists in the filaments. Scale bar, 20 nm. **e**, Three-dimensional density maps segmented from the original tomograms. Boxed densities are magnified views. **f**, Three-dimensional helical models were built that overlapped with the corresponding densities of PD- and MSA-associated fibrils, including a magnification of the central region. **g**, Helical models showing the periodicity of twisting of PD- or MSA-associated fibrils. Black arrows indicate the twist in the 3D model of the filament. **h**, Quantification of the periodic spacing (in nm) in many different fibrils derived from samples from patients with PD ($n = 3$) or patients with MSA ($n = 3$) samples. Each dot corresponds to a different fibril and data are mean \pm s.e.m. * $P < 0.05$ by one-way ANOVA followed by Tukey's multiple comparison test.

Circular dichroism spectroscopy showed that the secondary structure of α -syn aggregates in both PD and MSA predominantly comprises β -sheets (as illustrated by a negative peak at around 220 nm) (Fig. 3a). Analysis of the spectra indicates that MSA aggregates have a higher proportion of β -sheet structure than PD aggregates. Analogous results were obtained in the three samples from patients with PD and three samples from patients with MSA that were amplified from the brain rather than the CSF (Fig. 3b). To confirm these results using a different methodology, we used FTIR spectroscopy to estimate the secondary structures of α -syn aggregates in samples from a group of randomly selected patients with PD ($n = 10$) and patients with MSA ($n = 10$) (Fig. 3c). The MSA-derived aggregates showed a spectrum dominated by parallel β -sheet structure (peak at 1,640 cm $^{-1}$), whereas for PD-derived aggregates there was also another clear peak at around 1,652 cm $^{-1}$, which could be assigned to either α -helix- or random-coil-type structures (Fig. 3c).

To gain further insight into the structures of both species of α -syn, we performed cryo-electron tomography (cryo-ET) studies. Single-particle

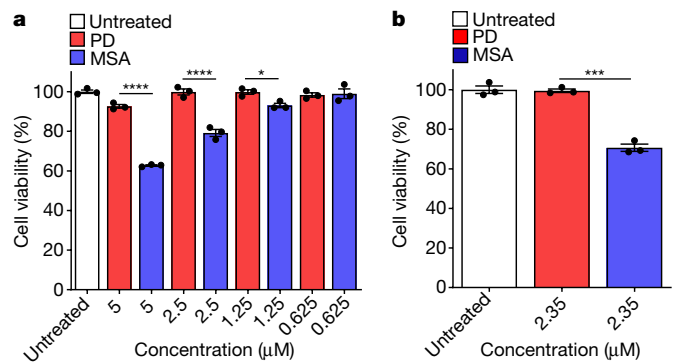


Fig. 4 | Cytotoxicity of amplified α -syn aggregates from the CSF of patients with PD or patients with MSA. **a**, RK13 cells (**a**) (10,000 cells), or neuronal precursor cells derived from human induced pluripotent stem cells generated as previously described²⁸ (**b**) (5,000 cells), were plated in a 96-well plate. After 24 h, cells were treated for 24 h for RK13 cells and 48 h for neuronal precursor cells with different concentrations of amplified α -syn fibrils from samples of CSF from patients with MSA or patients with PD. Cell viability was determined by MTT assay. Experiments were carried out in triplicate, each dot represents an individual replicate and data are mean \pm s.e.m. * $P < 0.05$, *** $P < 0.001$, **** $P < 0.0001$ by one-way ANOVA followed by Tukey's multiple comparison test.

cryo-electron microscopy (cryo-EM) has previously been used to determine the high-resolution structure of α -syn aggregates that were generated in vitro^{24,25}. Instead of taking single shots of two-dimensional (2D) images for a given area of a sample grid (as in single-particle cryo-EM), cryo-ET takes multiple shots in the same area by tilting the sample in a series of angles. A three-dimensional (3D) tomogram can be directly reconstructed from the series of tilts. To increase the contrast of the tomographic images, we negatively stained the fibrils amplified from the CSF of patients with PD or patients with MSA. We took 17- and 22-tilt series for PD and MSA samples, respectively (see 'Cryo-ET analysis and 3D reconstructions' in Methods for details). The tomograms (Fig. 3d, e) had enough contrast for us to determine that both fibrils were composed of two protofilaments that intertwine in a left-handed helix with a diameter of around 9 nm (see Extended Data Fig. 7 for more images of representative fibrils from three different patients). This is consistent with the high-resolution structure obtained by cryo-EM for full-length α -syn aggregates that were prepared in vitro²⁴. However, the lengths of fibril twists clearly varied between PD and MSA. On the basis of individual measurements of helical diameter and twist lengths, we were able to manually build helical models (Fig. 3f, g) guided by the segmented fibril densities (Fig. 3e). PMCA-derived α -syn aggregates from patients with PD were composed of long stretches of straight filaments with helical twists that generally ranged from 76.6 to 199 nm in length (Fig. 3g). By contrast, α -syn filaments from patients with MSA had shorter twists that mostly ranged from 46 to 105 nm in length (Fig. 3g). In accordance with this, measurements of periodic spacing indicated that the average twisting distance was significantly different between fibrils associated with PD and fibrils associated with MSA (65.2 \pm 3.8 nm (mean \pm s.e.m.) in MSA fibrils, $n = 104$ from 3 different patients; 108.5 \pm 6.1 nm in PD fibrils, $n = 104$ from 3 different patients) (Fig. 3h). These data indicate that the structures of α -syn aggregates derived from patients with PD and from patients with MSA are clearly different on the basis of their average periodicities of helical twists. Notably, previous studies using immuno-electron microscopy showed that non-amplified brain-derived α -syn filaments from patients with MSA are predominantly twisted²⁶, whereas those from patients with PD are mostly straight²⁷.

To explore whether aggregates derived from the CSF of patients with PD and patients with MSA have biological differences, we studied their toxicity in cell culture. For these experiments, we used a cell line that is often used in the prion field to study prion replication and

toxicity (RK13) (Fig. 4a), together with human neuronal precursor cells derived from induced pluripotent stem cells (Fig. 4b). Induced pluripotent stem cells and neuronal precursors were generated and characterized from fibroblasts obtained from a healthy individual, as previously described²⁸. We tested cytotoxicity by incubating cells with different concentrations of α -syn aggregates derived from the CSF of patients with PD or patients with MSA. MSA-derived aggregates showed highly significant toxicity in RK13 cells, even at concentrations of 1.25 μ M; by contrast, PD-derived aggregates began to show significant toxicity only at 5 μ M (Fig. 4a), indicating that MSA aggregates are more toxic than PD aggregates. A similar conclusion was obtained in the neuronal precursor cells that were derived from human induced pluripotent stem cells (Fig. 4b).

The prion-like behaviour of α -syn aggregates is a recently recognized principle that may have a central role in the pathological progression of various synucleinopathies^{29,30}. Indeed, the ability of α -syn aggregates to propagate their misfolded abnormalities enables the progressive spreading of damage from cell to cell^{3–5}. One of the tenets of the prion principle is that the misfolded protein can exist in different self-perpetuating conformational strains, which have the ability to faithfully template the misfolding of the normal monomeric protein in the abnormal-strain-specific conformation²⁹. Here we have shown that the prion principle can be used as an effective strategy to cyclically amplify the process of protein misfolding and thereby enable the detection of small amounts of α -syn aggregates in the CSF. Notably, we were able to distinguish—with high sensitivity and specificity—between samples from patients with two clinically similar synucleinopathies (PD and MSA). Moreover, we have shown that the α -syn aggregates present in the CSF of patients are representative of those that accumulate in the brain, indicating that the α -syn-PMCA assay can measure—non-invasively—the pathological species that are associated with different synucleinopathies. Our results demonstrate that α -syn aggregates exist as distinct conformational strains with different biochemical and structural properties, which will help to improve our understanding of the pathogenesis of these diseases. Furthermore, our study shows that patients with distinct synucleinopathies can be distinguished on the basis of the α -syn strain that is present in their CSF. These data may enable the development of a biochemical test for the specific diagnosis of different disorders that involve the misfolding of α -syn, with potential future applications in clinical trials and personalized medicine.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1984-7>.

- Goedert, M., Jakes, R. & Spillantini, M. G. The synucleinopathies: twenty years on. *J. Parkinsons Dis.* **7**, S51–S69 (2017).
- Wenning, G. K. et al. What clinical features are most useful to distinguish definite multiple system atrophy from Parkinson's disease? *J. Neurol. Neurosurg. Psychiatry* **68**, 434–440 (2000).

- Melki, R. Role of different alpha-synuclein strains in synucleinopathies, similarities with other neurodegenerative diseases. *J. Parkinsons Dis.* **5**, 217–227 (2015).
- Prusiner, S. B. et al. Evidence for α -synuclein prions causing multiple system atrophy in humans with parkinsonism. *Proc. Natl Acad. Sci. USA* **112**, E5308–E5317 (2015).
- Peng, C. et al. Cellular milieu imparts distinct pathological α -synuclein strains in α -synucleinopathies. *Nature* **557**, 558–563 (2018).
- Tarutani, A., Arai, T., Murayama, S., Hisanaga, S. I. & Hasegawa, M. Potent prion-like behaviors of pathogenic α -synuclein and evaluation of inactivation methods. *Acta Neuropathol. Commun.* **6**, 29 (2018).
- Shahnawaz, M. et al. Development of a biochemical diagnosis of Parkinson disease by detection of α -synuclein misfolded aggregates in cerebrospinal fluid. *JAMA Neurol.* **74**, 163–172 (2017).
- Kang, U. J. et al. Comparative study of cerebrospinal fluid α -synuclein seeding aggregation assays for diagnosis of Parkinson's disease. *Mov. Disord.* **34**, 536–544 (2019).
- Wood, S. J. et al. α -synuclein fibrillogenesis is nucleation-dependent. Implications for the pathogenesis of Parkinson's disease. *J. Biol. Chem.* **274**, 19509–19512 (1999).
- Volles, M. J. & Lansbury, P. T. Jr. Zeroing in on the pathogenic form of α -synuclein and its mechanism of neurotoxicity in Parkinson's disease. *Biochemistry* **42**, 7871–7878 (2003).
- El-Agnaf, O. M. et al. Detection of oligomeric forms of α -synuclein protein in human plasma as a potential biomarker for Parkinson's disease. *FASEB J.* **20**, 419–425 (2006).
- Tokuda, T. et al. Detection of elevated levels of α -synuclein oligomers in CSF from patients with Parkinson disease. *Neurology* **75**, 1766–1770 (2010).
- Herva, M. E. et al. Anti-amyloid compounds inhibit α -synuclein aggregation induced by protein misfolding cyclic amplification (PMCA). *J. Biol. Chem.* **289**, 11897–11905 (2014).
- Jung, B. C. et al. Amplification of distinct α -synuclein fibril conformers through protein misfolding cyclic amplification. *Exp. Mol. Med.* **49**, e314 (2017).
- Groveman, B. R. et al. Rapid and ultra-sensitive quantitation of disease-associated α -synuclein seeds in brain and cerebrospinal fluid by α syn RT-QuIC. *Acta Neuropathol. Commun.* **6**, 7 (2018).
- Fairfoul, G. et al. Alpha-synuclein RT-QuIC in the CSF of patients with alpha-synucleinopathies. *Ann. Clin. Transl. Neurol.* **3**, 812–818 (2016).
- Naiki, H., Higuchi, K., Hosokawa, M. & Takeda, T. Fluorometric determination of amyloid fibrils *in vitro* using the fluorescent dye, thioflavin T1. *Anal. Biochem.* **177**, 244–249 (1989).
- Sjöqvist, J. et al. Toward a molecular understanding of the detection of amyloid proteins with flexible conjugated oligothiophenes. *J. Phys. Chem. A* **118**, 9820–9827 (2014).
- Klingstedt, T. & Nilsson, K. P. Luminescent conjugated poly- and oligo-thiophenes: optical ligands for spectral assignment of a plethora of protein aggregates. *Biochem. Soc. Trans.* **40**, 704–710 (2012).
- Rasmussen, J. et al. Amyloid polymorphisms constitute distinct clouds of conformational variants in different etiological subtypes of Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **114**, 13018–13023 (2017).
- Sigurdson, C. J. et al. Prion strain discrimination using luminescent conjugated polymers. *Nat. Methods* **4**, 1023–1030 (2007).
- Bessen, R. A. & Marsh, R. F. Biochemical and physical properties of the prion protein from two strains of the transmissible mink encephalopathy agent. *J. Virol.* **66**, 2096–2101 (1992).
- Tuttle, M. D. et al. Solid-state NMR structure of a pathogenic fibril of full-length human α -synuclein. *Nat. Struct. Mol. Biol.* **23**, 409–415 (2016).
- Li, Y. et al. Amyloid fibril structure of α -synuclein determined by cryo-electron microscopy. *Cell Res.* **28**, 897–903 (2018).
- Guerrero-Ferreira, R. et al. Cryo-EM structure of alpha-synuclein fibrils. *eLife* **7**, e36402 (2018).
- Grazia Spillantini, M. et al. Filamentous α -synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. *Neurosci. Lett.* **251**, 205–208 (1998).
- Crowther, R. A., Daniel, S. E. & Goedert, M. Characterisation of isolated α -synuclein filaments from substantia nigra of Parkinson's disease brain. *Neurosci. Lett.* **292**, 128–130 (2000).
- Armijo, E. et al. Increased susceptibility to A β toxicity in neuronal cultures derived from familial Alzheimer's disease (PSEN1-A246E) induced pluripotent stem cells. *Neurosci. Lett.* **639**, 74–81 (2017).
- Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1332–1340 (2018).
- Olanow, C. W. & Prusiner, S. B. Is Parkinson's disease a prion disorder? *Proc. Natl Acad. Sci. USA* **106**, 12571–12572 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

Data reporting

No statistical methods were used to predetermine sample size, and the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Patient samples

CSF samples were obtained from 94 patients who were clinically diagnosed with PD, 75 patients who were diagnosed with MSA and 56 control individuals (people with other neurological diseases: epilepsy, cervical spondylosis, polyneuropathy, muscular dystrophy, viral myositis, myelopathy and hydrocephalus). Extended Data Table 1 displays a summary of the demographic characteristics of these patients. Most samples were collected at the Mayo Clinic, as indicated below. The clinical diagnoses of probable PD and MSA were made according to internationally standardized criteria, including the UK Brain Bank guidelines³¹. CSF samples were collected in the morning using polypropylene tubes following lumbar puncture at the L4/L5 or L3/L4 interspace with atraumatic needles after overnight fasting. The samples were centrifuged at 3,000g for 10 min at room temperature, aliquoted and stored at -80 °C until analysis. Blood cell (red and white) counts and glucose, protein and haemoglobin concentrations were determined as previously described⁷. The methods of CSF collection were approved by the institutional review boards at the study centres (Mayo Clinic and the University of Texas Health Science Center at Houston), and all study participants provided written informed consent.

Brain tissue from patients with PD and patients with MSA was obtained from the Banner Sun Health Research Institute. Control brain tissue was supplied by NDRI (National Human Tissue Resource Center). Frozen samples of frontal cortex were homogenized using a tissue grinder in 10% w/v ice-cold PBS (HyClone, SH30256.01) with complete protease inhibitor cocktail (Roche). The experiments with human tissue were performed following the universal precautions for working with human specimens and as directed by the Institutional Review Board of The University of Texas Health Science Center at Houston (HSC-MS-14-0608).

Expression and preparation of monomeric α -syn

The purification and characterization of monomeric α -syn was done as previously described⁷. In brief, the pET-21b plasmid carrying the coding DNA sequence for human α -syn containing a His-tag at the C terminus³² was overexpressed in BL21(DE3) pLysS (Invitrogen) *Escherichia coli* cells at 25 °C using 0.1 mM IPTG (isopropyl β -D-thiogalactoside) for 6 h. The bacterial pellets were lysed in 50 mM NaH₂PO₄ (pH 8.0), 300 mM NaCl, 10 mM imidazole, 1 mM PMSF, 0.1 mM tris-(2-carboxyethyl) phosphine (TCEP) and 1 mg ml⁻¹ lysozyme, followed by sonication on ice. The lysate was then centrifuged at 12,000g for 15 min at 4 °C, followed by ultracentrifugation at 100,000g for 30 min at 4 °C. The supernatant was filtered through a 0.45- μ m filter and loaded onto a nickel-affinity column (Nickel Sepharose Fast flow, GE Healthcare). Proteins were eluted using 250 mM imidazole and α -syn-containing fractions were dialysed overnight at 4 °C against PBS, pH 7.4. To remove any preformed seeds or aggregates, the protein solution was filtered through a 100-kDa cut-off filter (Amicon Ultra, Millipore), separated into small aliquots and stored at -80 °C until use. Protein concentration was determined by bicinchoninic acid (BCA) assay (Pierce). The purity of the protein was evaluated by silver staining.

α -syn-PMCA

The α -syn-PMCA (also known as α -syn-RT-QuIC) assay was performed as previously described⁷. In brief, samples of seed-free, monomeric α -syn at a concentration of 1 mg ml⁻¹ in 100 mM PIPES, pH 6.5 and 500 mM NaCl were placed in opaque 96-well plates (Costar, REF 3916) in the presence of 5 μ M ThT at a final volume of 200 μ l. For each test, we added

40 μ l of CSF from patients and controls or 40 μ l of brain homogenate (at a final concentration of 0.001%). Positive controls consisted of a well-documented and previously screened healthy CSF sample spiked with preformed α -syn oligomeric seeds. Samples were subjected to cyclic agitation (1 min at 500 rpm followed by 29 min without shaking) at 37 °C. The increase in ThT fluorescence was monitored at an excitation of 435 nm and emission of 485 nm, periodically, using a microplate spectrofluorometer Gemini-EM (Molecular Devices).

For serial rounds of amplification, an aliquot from the amplified material was diluted 100-fold into fresh α -syn monomer substrate and a new α -syn-PMCA assay was performed. This was repeated three consecutive times to obtain aggregates corresponding to the second, third and fourth rounds of amplification. The first round of amplification corresponds to the one initiated with the biological samples (CSF or brain homogenate).

Measurement of protein concentration in the aggregated product after amplification

Samples at the end of the PMCA reaction were centrifuged at 20,000g for 30 min at 4 °C. The resultant supernatants were carefully separated from the pellets. The amount of aggregated product was measured in all samples by three different procedures: (1) protein quantity in pellets was measured by silver staining after SDS-PAGE; (2) dot blot analysis of sedimented materials; and (3) BCA measurements of total protein content in the supernatant fraction. For SDS-PAGE, pellets were resuspended in PBS and separated on a 12% Bis-Tris gel and protein bands were visualized by silver staining as per the manufacturer's protocol. For dot blot analysis, 2 μ l of resuspended pellets was spotted onto nitrocellulose membranes (Amersham Biosciences) and air-dried for 30 min at room temperature. Blots were blocked with 5% w/v non-fat dry milk in Tris-buffered saline-Tween 20 (TBS-T) (20 mM Tris, pH 7.2, 150 mM NaCl and 0.05% (v/v) Tween 20) at room temperature for 2 h. After blocking, the membranes were probed with anti- α -syn antibody (BD Bioscience; 1:2,000) and anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibodies (1:5,000). The blots were visualized using enhanced chemiluminescence and a western blotting detection kit (Amersham Biosciences). Finally, the protein concentration in supernatants was determined using a BCA assay kit as per the manufacturer's recommendations.

Interaction of α -syn aggregates with thiophene-based ligands

A set of seven thiophene-based ligands (p-FTAA, h-FTAA, HS-68, HS-167, HS-169, HS-194 and HS-199) that have previously been shown to discriminate between different conformational strains composed of various proteins^{19,21,33} was used in this study. These compounds were synthesized and characterized as previously described^{19,33–36}, or as outlined below and in Extended Data Fig. 8 for compound HS-199. The stock solution for each compound was prepared in deionized water or DMSO at 1.5 mM. For our experiments, we diluted these stocks to reach a final concentration of 150 μ M. The excitation and emission wavelength range was different depending on the molecule, as previously described^{18,19}.

Synthesis and characterization of HS-199

A mixture of methyl 5'-bromo-[2,2'-bithiophene]-5-carboxylate (140 mg, 0.462 mM), (5-formylthiophen-2-yl)boronic acid (80 mg, 0.508 mM) (Extended Data Fig. 8), K₂CO₃ (192 mg, 1.39 mmol) in 1,4-dioxane/methanol (8: 2, 8 ml, degassed) and PEPPS-IPr (2 mol%) was heated to 80 °C for 30 min. After cooling to room temperature, the pH was adjusted to 4 by addition of 1 M HCl and the residue was extracted with DCM (3 \times 20 ml) and washed with water (3 \times 20 ml) and brine (30 ml). The combined organic phase was dried over MgSO₄ and the solvent was evaporated. The residue was subjected to column chromatography using CH₂Cl₂ followed by crystallization from DMF to give a trimer (Extended Data Fig. 8) as a yellow solid (115 mg, 74%).

A few drops of pyridine were added to a cold solution of this trimer (0.05 g, 0.150 mM) and the corresponding 2-methyl-3-alkylbenzothiazolium salt (Extended Data Fig. 8) (46 mg, 150 mM) in an anhydrous mixture of MeOH and THF (8:2). The mixture was refluxed until completion of the reaction (monitored by TLC, eluent: DCM/MeOH 1%). The solvent was evaporated *in vacuo* to provide a dark red solid, which was crystallized from MeOH. The red crystals were collected by filtration, washed with cold MeOH and dried in vacuum to afford HS-199 as a dark red solid (53 mg, 57%). Extended Data Figure 8 provides a summary of this reaction scheme.

The compound was characterized by infrared (IR) spectroscopy, nuclear magnetic resonance and mass spectrometry. IR (neat) 1,697, 7,594, 1,582, 1,525, 1,446, 1,421, 1,304, 1,245, 1,210, 1,165, 1,098, 1,054, 1,035, 926, 806, 786, 758 and 744 cm^{-1} . ^1H NMR (300 MHz, DMSO-*d*₆) δ 8.48–8.41 (m, 2H), 8.27 (d, *J* = 8.6 Hz, 1H), 7.97 (d, *J* = 3.9 Hz, 1H), 7.91–7.58 (m, 7H), 7.50 (d, *J* = 3.9 Hz, 1H), 4.92 (q, *J* = 7.0 Hz, 2H), 3.85 (s, 3H), 1.47 (t, *J* = 7.0 Hz, 3H). ^{13}C NMR (75 MHz, DMSO-*d*₆) δ 170.53, 161.43, 143.01, 142.17, 140.89, 140.83, 138.37, 137.10, 136.14, 136.01, 134.89, 131.30, 129.48, 28.26, 128.16, 127.79, 127.69, 126.66, 125.58, 124.38, 116.44, 111.21, 52.41, 44.29, 14.14. Matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF): *m/z* calculated for $\text{C}_{25}\text{H}_{20}\text{N}_2\text{O}_4$ (M+H)⁺: 495.0. Found: 495.0.

Protease digestion and epitope mapping

Samples containing α -syn aggregates amplified by PMCA were treated with different concentrations of proteinase K at 37 °C for 1 h. The reaction was stopped by heating the sample in NuPAGE LDS buffer at 95 °C for 10 min. The digested products were resolved by 12% Bis-Tris gels (Invitrogen). Proteins were electrophoretically transferred to nitrocellulose membranes (Amersham Biosciences). Membranes were blocked with 5% w/v non-fat dry milk in PBS–Tween 20 (PBS (Hyclone SH.30258.02, pH 7.2, 0.1% (v/v) Tween 20) at room temperature for 1 h. After blocking, the membranes were probed with the following antibodies against α -syn: N-19 (Santa Cruz), which recognizes the N terminus (residues 1–50) of α -syn; anti- α -syn clone 42 (BD Biosciences), which is raised against the middle region (residues 15–123) of the protein; and 211 (Santa Cruz), which is reactive against the C-terminal region (residues 121–125) of α -syn. The blots were developed using ECL prime detection western blotting reagents (Amersham Biosciences).

Circular dichroism

Solutions containing around 35 μM of α -syn aggregates amplified by α -syn-PMCA were used for these studies. Circular dichroism spectra were recorded at room temperature using a JASCO J815 spectropolarimeter, with 1-mm path-length cuvette. Circular dichroism data were collected at 0.1-nm resolution and at a scan speed of 200 nm min^{-1} . The portion of the circular dichroism spectrum between 250 and 350 nm was fitted with a quadratic function and the baseline of the whole spectrum was calculated using the function. Then the calculated baseline was subtracted from the circular dichroism spectrum to obtain the baseline-corrected circular dichroism spectrum.

FTIR spectroscopy

FTIR experiments were conducted using an FT/IR-4100 spectrometer from JASCO. The product of α -syn-PMCA (5 μl) was placed on the top of a diamond PRO450-S attenuated total reflectance unit (JASCO) adapted to the FT/IR-4100 system. The system parameters included a resolution of 4.0 cm^{-1} and an accumulation of 80 scans per sample. The data were processed using cosine apodization and Mertz phase correction. The data were also corrected for attenuated total reflectance and carbon dioxide vapour absorption.

Cryo-ET analysis and 3D reconstructions

The product of α -syn-PMCA (after 2 rounds of amplification from CSF samples from patients with PD or patients with MSA) was sedimented

at 20,000g for 30 min at 4 °C, resuspended in 100 mM PIPES, pH 6.5 and 500 mM NaCl, diluted 10-fold in deionized water and loaded onto Formvar/Carbon Copper grids. Samples were negatively stained with 2% uranyl acetate and rapidly frozen in liquid ethane, using a gravity-driven plunger apparatus. Materials were imaged at -170 °C using a Polara G2 electron microscope (FEI) equipped with a field-emission gun and a direct-detection device (Gatan K2 Summit). The microscope was operated at 300 kV with a magnification of $\times 15,500$. We used SerialEM³⁷ to collect tomographic tilt series at a defocus of around 6 μm , with cumulative doses of around 200 e^- per \AA^2 . For each dataset, 35 image stacks were collected in a range from -51° to $+51^\circ$, using increments of 3° . Each stack contained about 10 images, which were first aligned using MotionCor2³⁸. The tomograms were reconstructed using IMOD software³⁹ and were further processed by EMAN software⁴⁰.

The helical models were manually built based on individual fibril density. The twist lengths of fibrils vary from one to another. We were not able to perform subtomogram averaging owing to the heterogeneous nature of the fibrils. The single fibril density used for modelling has very limited resolution. Its density in the *z*-direction is elongated because of a missing wedge issue (no high-angle tilt images). We rely on the helical parameters (diameter and the twist lengths) to build a helical model, which can be directly measured from the centre slice of the tomogram in the *x–y* plane. Instead of creating a mathematical helical model on the basis of the two parameters, we manually built the helical model by tracing the filament density using Chimera software⁴¹. Model dots were placed along the densities followed by manual adjustment of the dot positions to make the model shape helix-like under the restriction of cryo-ET density. Although density restriction was applied, the constructed pseudo-helical model does not completely fit into the noisy and distorted fibril density.

Cytotoxicity assays

RK13 cells (rabbit kidney cell line, ATCC CCL-37) were grown in DMEM medium supplemented with 10% FBS $1\times$ GLUTamax, $1\times$ MEM and 1 mM sodium pyruvate. For toxicity, 10,000 cells were plated without antibiotic in a 96-well plate and incubated at 37 °C for 24 h. Neuronal precursors derived from human induced pluripotent stem cells were generated and characterized as previously described²⁸. These cells were maintained in neural precursor expansion medium (NPEM) as previously described. Approximately 5,000 cells were plated per well in a 96-well plate, pre-coated with Geltrex LDEV-free reduced growth factor basement membrane matrix-treated dishes (1:100, Invitrogen) and incubated at 37 °C for 24 h. After 24 h, cells were treated for either 24 h (RK13 cells) or 48 h (neuronal precursors) with different concentration(s) of amplified α -syn fibrils originating from CSF samples from patients with MSA and patients with PD. Cell viability was determined by the MTT assay, following the manufacturer's protocol.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated and/or analysed during this study are included in the Article, Supplementary Fig. 1 (uncropped blots) and the Source Data files for Figs. 1, 3, 4 and Extended Data Figs. 1–3. Any additional information required are available from the corresponding author on reasonable request.

31. Tolosa, E., Wenning, G. & Poewe, W. The diagnosis of Parkinson's disease. *Lancet Neurol.* **5**, 75–86 (2006).

32. Roostaei, A., Beaudoin, S., Staskevicius, A. & Roucou, X. Aggregation and neurotoxicity of recombinant α -synuclein aggregates initiated by dimerization. *Mol. Neurodegener.* **8**, 5 (2013).

33. Åslund, A. et al. Novel pentameric thiophene derivatives for in vitro and in vivo optical imaging of a plethora of protein aggregates in cerebral amyloidoses. *ACS Chem. Biol.* **4**, 673–684 (2009).
34. Klingstedt, T. et al. Distinct spacing between anionic groups: an essential chemical determinant for achieving thiophene-based ligands to distinguish β -amyloid or tau polymorphic aggregates. *Chemistry* **21**, 9072–9082 (2015).
35. Shirani, H. et al. Synthesis of thiophene-based optical ligands that selectively detect tau pathology in Alzheimer's disease. *Chemistry* **23**, 17127–17135 (2017).
36. Shirani, H. et al. A palette of fluorescent thiophene-based ligands for the identification of protein aggregates. *Chemistry* **21**, 15133–15137 (2015).
37. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
38. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
39. Mastronarde, D. N. & Held, S. R. Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102–113 (2017).
40. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
41. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements This study was funded in part by grants from the Michael J. Fox Foundation for Parkinson's disease (to C.S. and S.P.); NIH (R01AG055053, R01AG061069) and Department of Defense (to C.S.); NIH (P01NS44233, U54NS065736, K23NS075141, R01FD004789, R01NS092625), Department of Defense and Mayo Funds (to P.A.L.); R01NS094535 (to A.-L.T.); and the Swedish Research Council (2016-00748 to H.S. and K.P.R.N.). We are grateful to the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for the provision of brain tissue. We also thank N. P. Rocha for providing CSF samples, I. Moreno-Gonzalez for helping with the preparation and characterization of brain homogenate and T. Eckland for editing the manuscript.

Author contributions C.S. and M.S. conceived and designed the experiments and analysed the data, with important contributions from A.M. and S.P. for some of the experiments; M.S. performed all PMCA assays, analysed data and prepared figures; A.M. performed FTIR assays, analysed data and prepared figures; S.P. performed assays with thiophene-based ligands, analysed data and prepared figures; N.M. performed all protease-resistance and epitope-mapping experiments, prepared figures and performed the sedimentation studies; P.R. purified the recombinant α -syn for the experiments; X.L. and B.H. performed cryo-ET, constructed models and measured pitch lengths; X.L. discovered the key difference in the pitch length between PD and MSA fibrils; C.S. and A.M. analysed the cryo-ET data and prepared figures; A.M., G.W. and A.L.-T. performed circular dichroism spectroscopy, analysed data and prepared figures; M.S. and A.M. performed cytotoxicity assays, analysed data and prepared figures; H.S. and K.P.R.N. provided thiophene-based ligands and experimental support for their use; A.S., W.S. and P.A.L. provided most of the CSF samples and clinical data; C.S. wrote the manuscript with input from all co-authors.

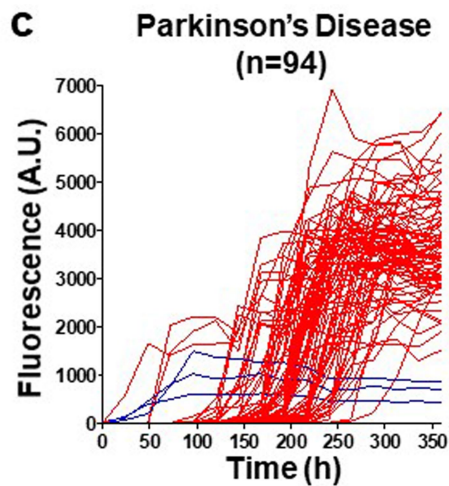
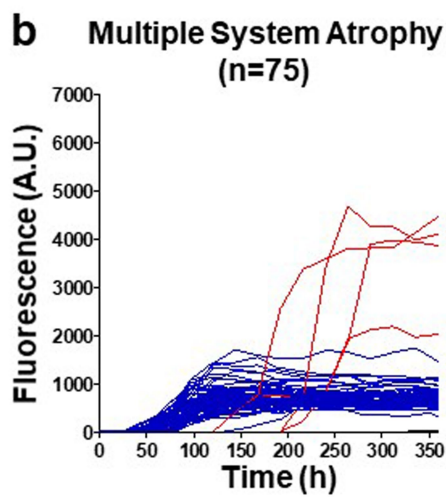
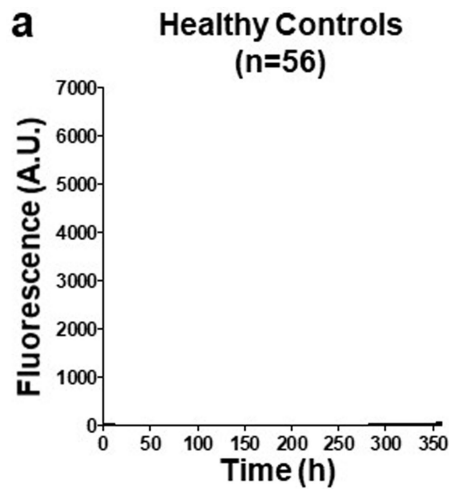
Competing interests C.S. and M.S. are inventors on patent applications (US20160077111, WO2016040905, EP3191599A1, US20160077112 and WO2016040907) for the use of PMCA technology for high-sensitive detection of α -syn aggregates in patients affected by synucleinopathies. These applications were filed by the University of Texas Health Science Center at Houston and Amprion Inc. C.S. is an inventor on several patents related to PMCA technology and is a Founder, Chief Scientific Officer and Member of the Board of Directors of Amprion Inc, a biotechnology company that focuses on the commercial use of PMCA (RT-QuIC) for high-sensitivity detection of misfolded protein aggregates that are implicated in a variety of neurodegenerative diseases. The University of Texas Health Science Center at Houston owns some patent applications related to the PMCA (RT-QuIC) technology that have been licensed to Amprion Inc.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1984-7>.

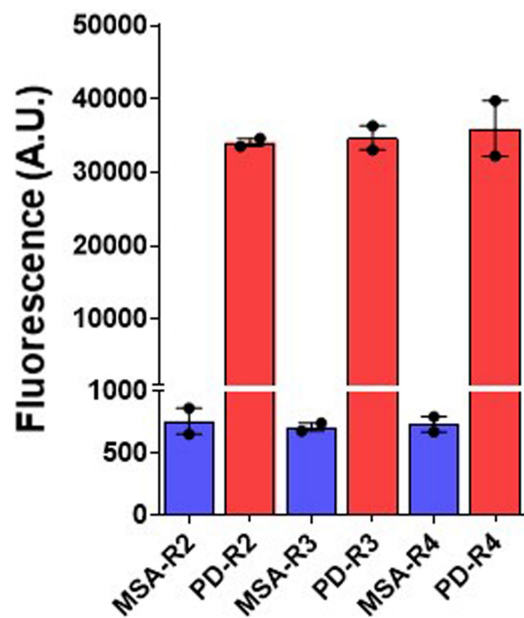
Correspondence and requests for materials should be addressed to C.S.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

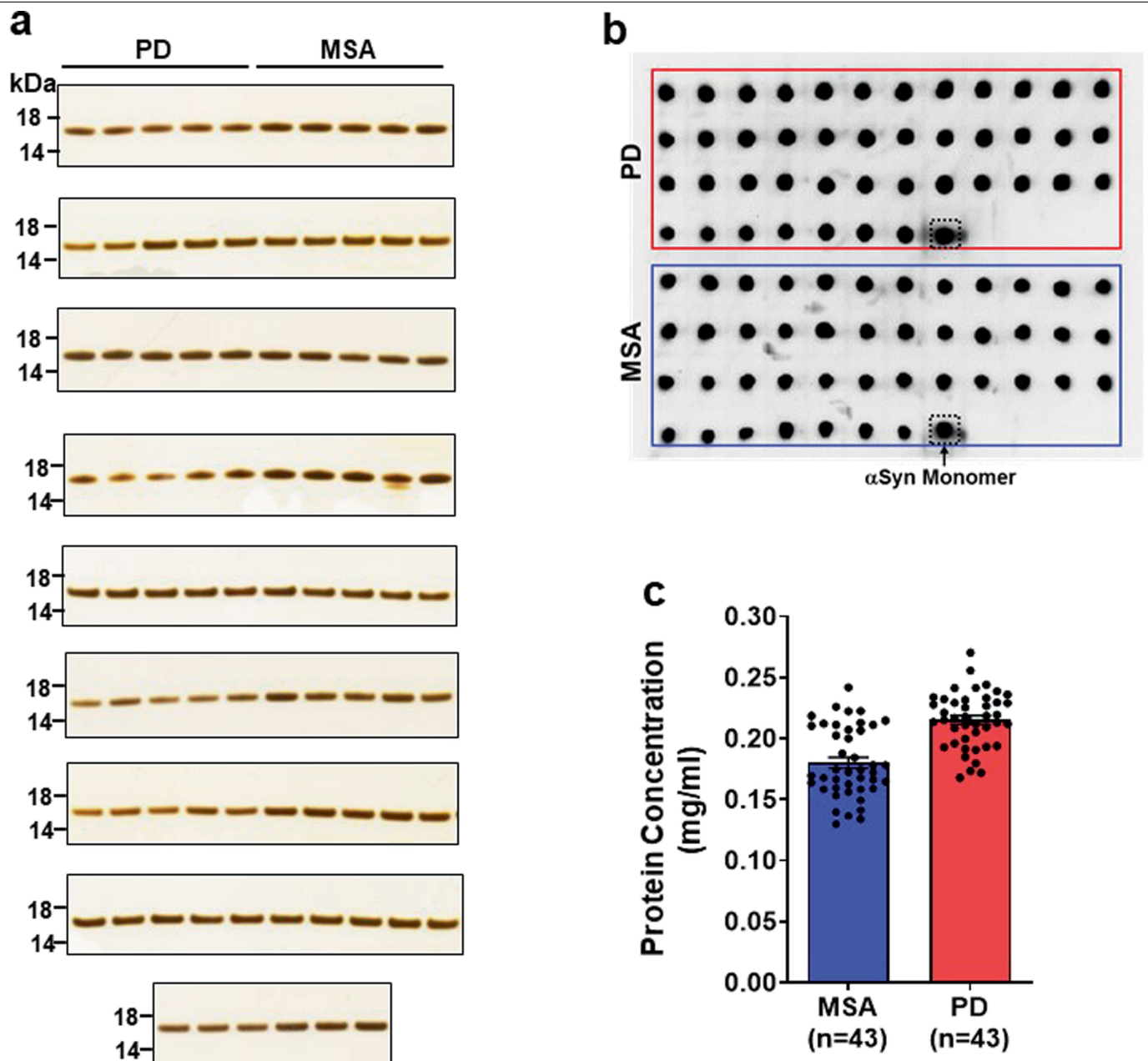


Extended Data Fig. 1 | Kinetics of α -syn aggregation in the presence of CSF from patients with PD, patients with MSA or healthy control individuals.

a–c, Individual α -syn aggregation curves are shown in the presence of CSF samples (40 μ l) from all study participants, including healthy controls (**a**; $n = 56$), patients with MSA (**b**; $n = 75$) and patients with PD (**c**; $n = 94$). The α -syn-PMCA assay was started by adding α -syn monomers (1 mg ml⁻¹) and ThT (5 μ M) to 100 mM PIPES, pH 6.5 containing 500 mM NaCl. The plate was incubated at 37 °C with intermittent shaking for 1 min every 30 min at 500 rpm. The extent of aggregation was monitored using a fluorometer to measure ThT fluorescence, with an excitation of 435 nm and emission of 485 nm. The colours represent the expected aggregation curves for patients with PD (red), patients with MSA (blue) and healthy controls (black), regardless of clinical diagnosis.

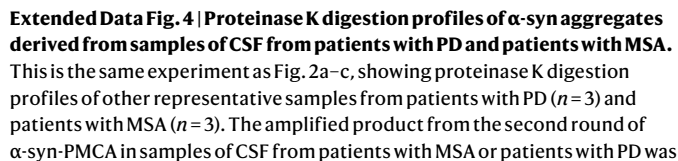


Extended Data Fig. 2 | Serial propagation of α -syn aggregates derived from patients with MSA and patients with PD. For serial propagation of α -syn aggregates, an aliquot of the final product of the first α -syn-PMCA reaction (starting from CSF samples) was diluted 100-fold into a solution containing fresh α -syn monomers (1 mg ml^{-1}). A second round of amplification was done in the same buffer (100 mM PIPES, pH 6.5 containing 500 mM NaCl) at 37°C with intermittent shaking for 1 min every 30 min at 500 rpm. The extent of aggregation was monitored by the increase in ThT fluorescence. The maximum fluorescence value at the plateau of aggregation was recorded and plotted in the graph as the second round of amplification (R2). Similarly, the third and fourth rounds of amplification (R3 and R4) were performed by diluting the product 100-fold on amplification each time into fresh α -syn monomer substrate and repeating the α -syn-PMCA assay. The results shown are from one patient with PD and one patient with MSA. The experiment was carried out in duplicate, each dot represents an individual technical replicate and data are mean \pm s.e.m.

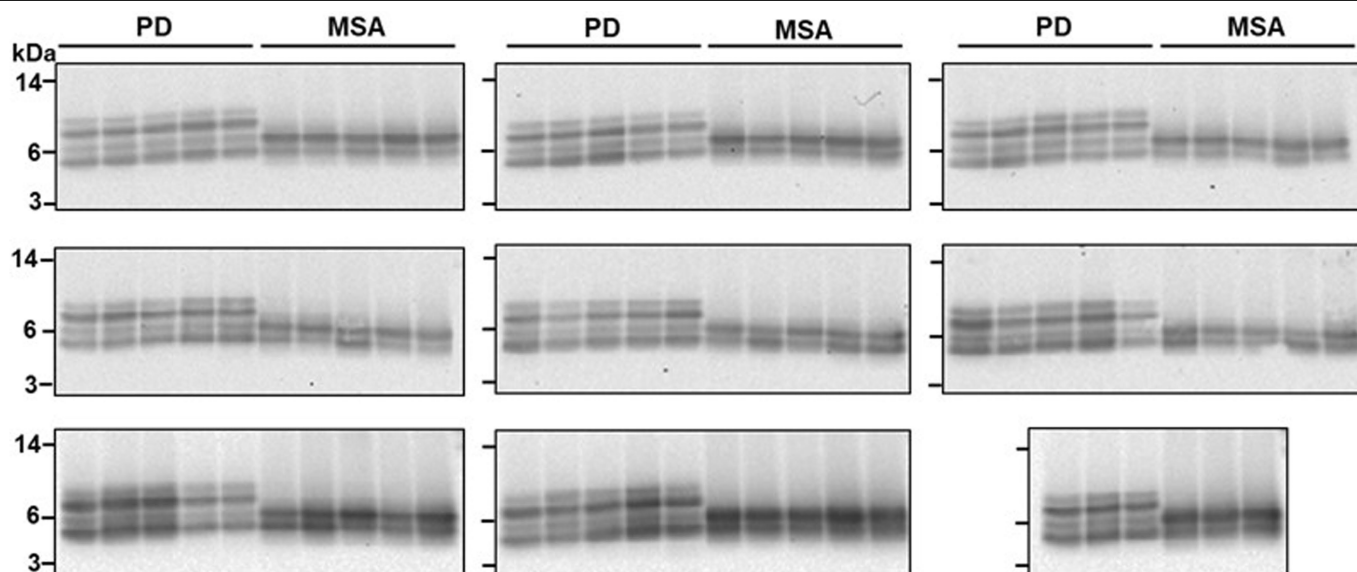


Extended Data Fig. 3 | Analyses of the quantity of α -syn aggregates after amplification from patients with MSA and patients with PD by sedimentation assay. Aggregates of α -syn that were obtained after two rounds of α -syn-PMCA amplification (starting from CSF samples from patients with MSA ($n = 43$) and patients with PD ($n = 43$)) were centrifuged at 20,000g for 30 min. **a**, The resultant pellets were separated on a 12% Bis-Tris gel, and protein bands were visualized by silver staining as per the manufacturer's protocol. Molecular weight markers (kDa) are indicated on the left of the gel. **b**, Resuspended pellets (2 μ l) were spotted onto nitrocellulose membranes and air-dried for 30 min at room temperature. After blocking with 5% w/v non-fat

dry milk at room temperature for 2 h, membranes were probed with an anti- α -syn antibody (BD Bioscience, 1:2,000) and anti-rabbit HRP-conjugated secondary antibodies (1:5,000). The blots were visualized using enhanced chemiluminescence and a western blotting detection kit. The dot blot shows each of the 86 samples ($n = 43$, PD; $n = 43$, MSA) and a positive control using non-aggregated α -syn monomer (dotted box). The results are representative of two independent experiments with similar results. **c**, Protein concentration in the supernatants was determined by a BCA assay kit as per the manufacturer's instructions. Each dot represents an individual sample ($n = 43$, PD; $n = 43$, MSA) in each disease group and data are mean \pm s.e.m.

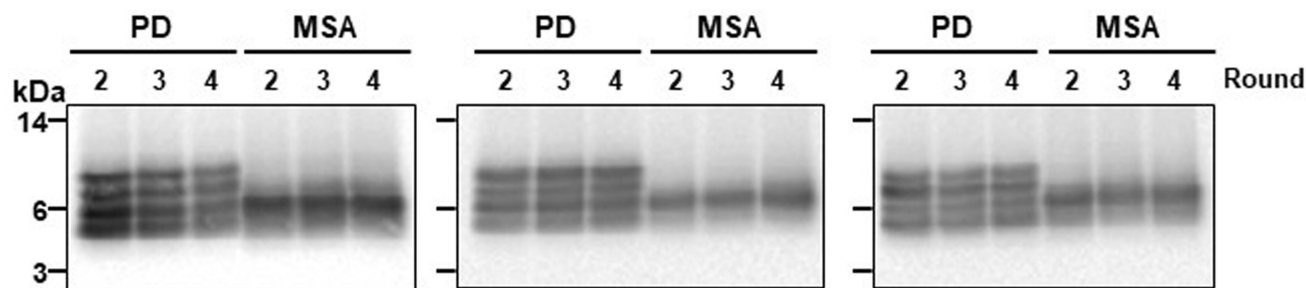


incubated either without (–) or in the presence of increasing concentrations of proteinase K (0.001, 0.01, 0.1 and 1 mg ml^{–1}) at 37 °C for 1 h. Proteins were separated on a 12% Bis-Tris gel and immunoblotted with the same antibodies as in Fig. 2 (SCN-19 (top), BD anti- α -syn clone 42 (middle) and SC 211 (bottom)). Each blot represents an individual sample. Molecular weight markers (kDa) are indicated on the left of the blot.



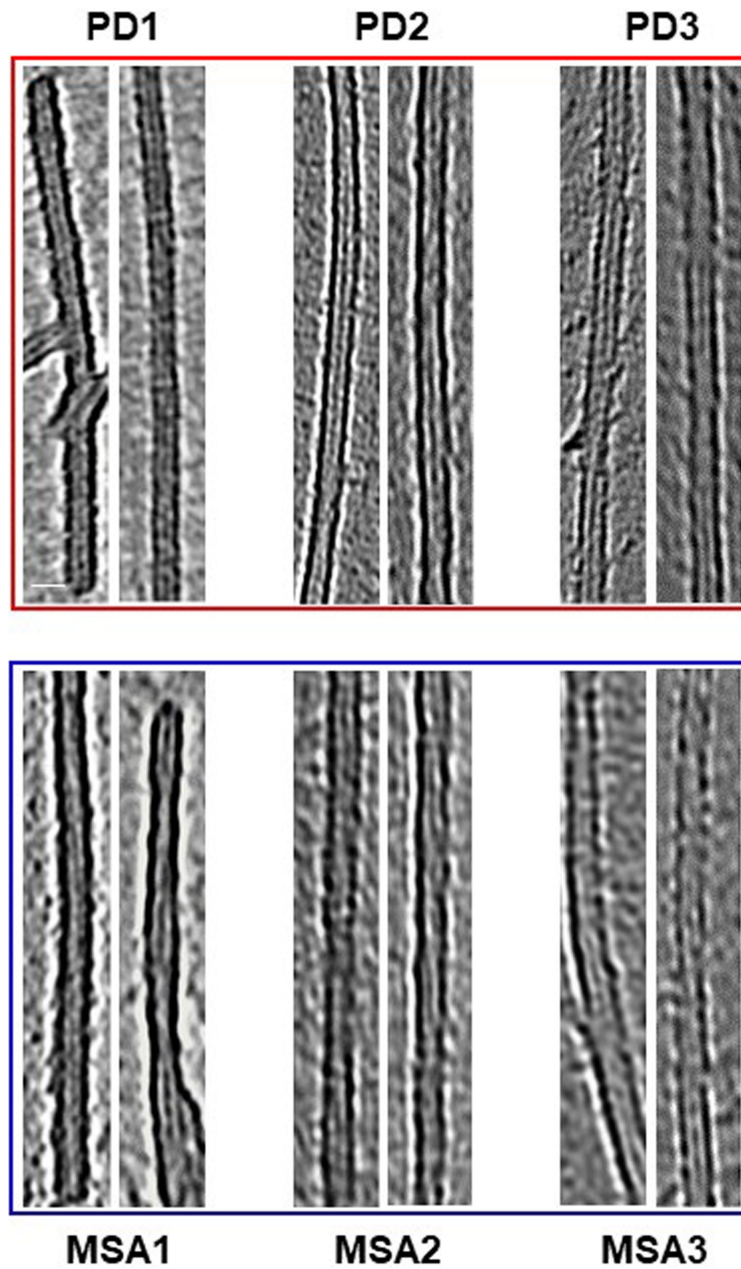
Extended Data Fig. 5 | Proteinase K digestion profiles of α -syn aggregates derived from samples of CSF from all 43 patients with PD and 43 patients with MSA. This is the same experiment as Fig. 2d, showing proteinase K digestion profiles of all 86 ($n = 43$, PD; $n = 43$, MSA) biologically independent samples analysed. Aliquots of the product of the second round of the α -syn-

PMCA assay were treated with proteinase K (1 mg ml^{-1}) at 37°C for 1 h. Proteins were separated on a 12% Bis-Tris gel and immunoblotted with the BD anti- α -syn clone 42 antibody. Molecular weight markers (kDa) are indicated on the left of the blot. The third blot on the top row is the same as that shown in Fig. 2d.



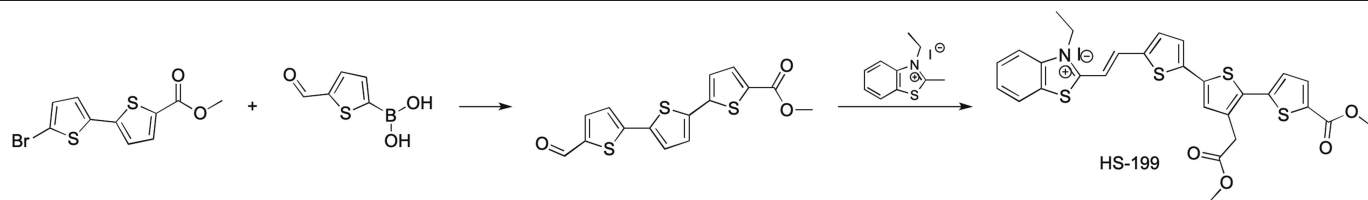
Extended Data Fig. 6 | Proteinase K digestion profiles of α -syn aggregates after several rounds of α -syn-PMCA. This is the same experiment as Fig. 2e, showing the results obtained with samples from different patients with PD ($n=3$) and patients with MSA ($n=3$). The first round corresponds to direct amplification from the CSF of the patients. For the second round of amplification, aggregates produced in the first round were diluted 100-fold into fresh α -syn monomer substrate and a new round of α -syn-PMCA was

performed. The assay was then repeated for the third and fourth rounds using amplified α -syn aggregates (1%) from the previous round. Amplified aggregates were treated with proteinase K (1 mg ml^{-1}) for 1 h and proteins were separated on a 12% Bis-Tris gel and immunoblotted with the BD anti- α -syn clone 42 antibody. Molecular weight markers (kDa) are indicated on the left of the blot.



Extended Data Fig. 7 | Electron microscopy images of PD-associated fibrils and MSA-associated fibrils. Representative images of fibrils produced after two rounds of α -syn-PMCA in samples from different patients with PD ($n = 3$)

and patients with MSA ($n = 3$). The negative-stained fibrils were imaged with a 300 kV electron microscope. Scale bar, 10 nm (applies to all of the images).



Extended Data Fig. 8 | Reaction scheme for the chemical synthesis of HS-199.

HS-199 was synthesized by mixing 0.462 mM methyl 5'-bromo-[2,2'-bithiophene]-5-carboxylate with 0.508 mM (5-formylthiophen-2-yl)boronic

acid, K_2CO_3 (1.39 mmol) in 1,4-dioxane/methanol (8:2, 8 mL/mM, degassed) and PEPPS-IPr (2 mol %).

Extended Data Table 1 | Number of samples and basic demographic information for all study participants

	Healthy Controls (HC)		Multiple System Atrophy (MSA)		Parkinson's Disease (PD)	
Number of samples	56		75		94	
Age (Mean \pm SD)	60.14 \pm 10.18		59.67 \pm 6.9		66.78 \pm 7.91	
Sex (M,F)	M = 28	F = 28	M = 56	F = 19	M = 59	F = 36
Disease Duration (Mean \pm SD)	N/A		3.847 \pm 2.6		8.13 \pm 4.7	

Age and disease duration are given in years.

Extended Data Table 2 | Demographic information for the 43 patients with PD and 43 patients with MSA whose CSF samples were used to characterize amplified α -syn aggregates in detail

Parkinson's Disease (PD)				Multiple System Atrophy (MSA)			
S/No	Clinical Diagnosis	Age/Sex	Disease Duration	S/No	Clinical Diagnosis	Age/Sex	Disease Duration
1	PD	71/F	11	1	MSA-P	61/M	2
2	PD	61/M	8	2	MSA-C	60/M	6
3	PD	59/F	12	3	MSA-C	51/M	5
4	PD	63/F	7	4	MSA-C	66/F	3
5	PD	66/F	8	5	MSA-C	58/M	7
6	PD	62/M	8	6	MSA-C	53/M	2
7	PD	72/M	17	7	MSA-C	65/F	2
8	PD	69/M	12	8	MSA-P	55/M	3
9	PD	67/M	15	9	MSA-P	64/M	1
10	PD	58/F	8	10	MSA-C	60/M	6
11	PD	66/M	6	11	MSA-C	63/M	1
12	PD	74/F	6	12	MSA-C	51/M	2.5
13	PD	73/F	5	13	MSA-P	59/F	2
14	PD	61/F	9	14	MSA-C	54/M	6
15	PD	69/M	5	15	MSA-P	49/F	3
16	PD	65/M	7	16	MSA-C	55/M	3
17	PD	69/F	11	17	MSA-C	68/M	5
18	PD	65/M	10	18	MSA-C	56/F	1.5
19	PD	68/M	N/A	19	MSA-C	57/F	2.5
20	PD	57/F	N/A	20	MSA-C	59/M	1
21	PD	71/F	N/A	21	MSA-C	49/M	1
22	PD	68/M	N/A	22	MSA-C	50/M	2
23	PD	79/M	N/A	23	MSA-P	54/M	5
24	PD	68/M	21	24	MSA-P	64/M	3
25	PD	70/M	1	25	MSA-C	68/F	6
26	PD	62/M	3	26	MSA-C	60/M	2
27	PD	63/M	10	27	MSA-C	61/M	1
28	PD	74/M	17	28	MSA-C	62/M	2
29	PD	69/M	6	29	MSA-P	53/M	4
30	PD	66/M	5	30	MSA-P	54/M	5
31	PD	73/F	15	31	MSA-C	61/M	8
32	PD	54/M	6	32	MSA-C	66/F	3
33	PD	71/M	5	33	MSA-C	54/M	4
34	PD	72/M	5	34	MSA-C	62/M	9
35	PD	74/M	5	35	MSA-C	48/F	5
36	PD	59/M	15	36	MSA-P	54/F	2
37	PD	53/F	4	37	MSA-P	60/F	2
38	PD	73/M	N/A	38	MSA-P	73/F	10
39	PD	69/F	N/A	39	MSA-P	57/F	2.5
40	PD	47/M	N/A	40	MSA-C	68/F	3.5
41	PD	59/M	N/A	41	MSA-C	52/M	12
42	PD	61/F	24	42	MSA-C	58/M	5
43	PD	68/M	12	43	MSA-C	54/M	3

Age and disease duration are given in years. MSA-C, MSA with cerebellar ataxia; MSA-P, MSA with Parkinsonism.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

SerialEM was used for collecting cryo-ET data

Data analysis

Statistic analysis were done by Graph Pad Prism, version 7. Cryo-ET data was analyzed by IMOD and EMAN software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Authors will make raw data available upon request, without any restriction.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used the maximum number of samples that we were able to obtain, which was much higher than needed according power analysis
Data exclusions	No data was excluded from the study
Replication	Results were replicated in independent experiments as described in the text. Also, every experiment included several replicates as describe in the figure legends. All experiments to replicate the results were successful.
Randomization	All samples were used without any randomization, but based on the maximum number of samples available. The demographic information is provided in Tables. Results obtained are not different by gender, race or any variable other than clinical diagnosis.
Blinding	Except for experiments to optimize the methodologies, experimenter was always blinded to the sample origin.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	All antibodies used are from commercial source. The name of the antibody, manufacturer, dilution, etc are listed in the attached page
Validation	The antibodies are extensively used in the field and are well validated by the manufacturer as described in the attached document.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Cell lines are commercially available. The name and precedence is described in the text as well as in the attached document.
Authentication	Cell lines have been validated by the vendor as stated in the attached document.
Mycoplasma contamination	We routinely test cell lines for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	We did not use any misidentified cell line

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	CSF and brain samples from individuals with clinical diagnosis of Parkinson's disease, multiple system atrophy and controls were used. The demographic characteristics of these patients, including diagnosis, is included in tables as extended data.
Recruitment	Patients were recruited at Mayo Clinic purely based on clinical diagnosis. No especial requirement for recruitment that could bias the sample population.
Ethics oversight	Human subject protocol was reviewed and approved at the University of Texas Health Science Center at Houston and Mayo Clinic

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)
All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	<i>Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.</i>
Study protocol	<i>Note where the full trial protocol can be accessed OR if not available, explain why.</i>
Data collection	<i>Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.</i>
Outcomes	<i>Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.</i>

Antibodies

- 1) Polyclonal goat anti- α/β -Synuclein (N-19), Cat. No. sc-7012, Santa Cruz, Dilution 1:2000, Lot. No. K1213.
- 2) Monoclonal mouse anti- α -Synuclein (211), Cat. No. sc-12767, Santa Cruz, Dilution 1:4000, Lot. No. B1518.
- 3) Monoclonal mouse anti- α -Synuclein (clone 42), Cat. No. 610787, BD Biosciences, Dot Blot Dilution 1:2000, Western Blot Dilution 1:5000, Lot. No. 7243572
- 4) Sheep anti-mouse IgG conjugated to horseradish peroxidase (HRP), Cat. No. A5906, Sigma Aldrich, Dot Blot Dilution 1:5000, Western Blot Dilution 1:10000, Lot. No. SLBT9505
- 5) Donkey anti-goat IgG conjugated to horseradish peroxidase (HRP), Cat. No. A15999, Invitrogen, Dilution 1:10000

Validation

All antibodies were validated according to the manufacturer's instruction and as described in the literature

- 1) Commercial antibody, validation data available on manufacturer's website. Ref: Sharma N, et al. 2001 Acta Neuropathologica 102(4):329-34.
- 2) Commercial antibody, validation data available on manufacturer's website. Ref: Zunke F, Moise AC, et al. 2018 Neuron 97(1):92-107.
- 3) Commercial antibody, validation data available on manufacturer's website. Ref: Liu Y, Fallon L, et al. 2002 Cell 111(2):209-18.
- 4) Commercial antibody, validation data available on manufacturer's website. Ref: Trowitzsch S, Viola C, et al. 2015 Nature Communications 6: 6011.
- 5) Validated by Manufacturer, see Certificate of Analysis on website

Bacteria

BL21 Star™ (DE3)pLysS One Shot™ Chemically Competent *E. coli*

Cat. No. 44-0054

Lot. No. 1437577

Invitrogen

Eukaryotic cell lines

Cell line source: RK13 (ATCC® CCL37™) cell lines were purchased from ATCC.

Authentication: ATCC provided certificate of analysis for RK13 (ATCC® CCL37™) cell line.

Cell line source: Neural precursor cells were generated from human iPSC cells reprogrammed from skin fibroblasts obtained from a 66 year old female (cell number: AG08517) purchased from (Coriell, Camden, NJ, USA). The procedure for reprogramming, differentiation and growing of cells is described in detail in our previous publication (Armijo, E. et al. Neurosci. Lett 639, 74-81, 2017).

Authentication: Cells were fully characterized by various procedures as described in detail in our previous publication (Armijo, E. et al. Neurosci. Lett 639, 74-81, 2017).

Mycoplasma contamination: All cell lines were tested routinely negative for Mycoplasma contamination by PCR.

Commonly misidentified lines (See ICLAC register): No commonly misidentified lines were used.

Live-animal imaging of native haematopoietic stem and progenitor cells

<https://doi.org/10.1038/s41586-020-1971-z>

Received: 7 June 2018

Accepted: 6 December 2019

Published online: 5 February 2020

Constantina Christodoulou^{1,2,13,14}, Joel A. Spencer^{3,4,5,6,14}, Shu-Chi A. Yeh^{3,4,14}, Raphaël Turcotte^{3,4}, Konstantinos D. Kokkalis⁷, Riccardo Panero⁸, Azucena Ramos^{1,2}, Guoji Guo⁹, Negar Seyedhassantehrani⁶, Tatiana V. Esipova^{10,11}, Sergei A. Vinogradov^{10,11}, Sarah Rudzinskas¹², Yi Zhang¹², Archibald S. Perkins¹², Stuart H. Orkin⁹, Raffaele A. Calogero⁸, Timm Schroeder⁷, Charles P. Lin^{3,4*} & Fernando D. Camargo^{1,2*}

The biology of haematopoietic stem cells (HSCs) has predominantly been studied under transplantation conditions^{1,2}. It has been particularly challenging to study dynamic HSC behaviour, given that the visualization of HSCs in the native niche in live animals has not, to our knowledge, been achieved. Here we describe a dual genetic strategy in mice that restricts reporter labelling to a subset of the most quiescent long-term HSCs (LT-HSCs) and that is compatible with current intravital imaging approaches in the calvarial bone marrow^{3–5}. We show that this subset of LT-HSCs resides close to both sinusoidal blood vessels and the endosteal surface. By contrast, multipotent progenitor cells (MPPs) show greater variation in distance from the endosteum and are more likely to be associated with transition zone vessels. LT-HSCs are not found in bone marrow niches with the deepest hypoxia and instead are found in hypoxic environments similar to those of MPPs. In vivo time-lapse imaging revealed that LT-HSCs at steady-state show limited motility. Activated LT-HSCs show heterogeneous responses, with some cells becoming highly motile and a fraction of HSCs expanding clonally within spatially restricted domains. These domains have defined characteristics, as HSC expansion is found almost exclusively in a subset of bone marrow cavities with bone-remodelling activity. By contrast, cavities with low bone-resorbing activity do not harbour expanding HSCs. These findings point to previously unknown heterogeneity within the bone marrow microenvironment, imposed by the stages of bone turnover. Our approach enables the direct visualization of HSC behaviours and dissection of heterogeneity in HSC niches.

At present, tracking of HSCs in live animals requires transplantation of the HSCs to be imaged, typically in the calvarium of an irradiated recipient whose bone marrow microenvironment has been severely altered^{4,6}. Therefore, although engraftment biology can be studied in these models, the behaviour of stem cells and progenitors is likely to differ from that seen in the unperturbed state^{1,2,4}. The recent description of HSC-reporter lines in mice has facilitated the identification of these cells in bone sections and after tissue clearing; nevertheless, these reporters are still not fully HSC-specific and require the use of additional markers^{7–9}. Despite these advances, there is still considerable uncertainty about the exact localization of HSC and progenitor cells. Even less is known about the nature of distinct niches that support HSC proliferation or maintain HSC quiescence⁷.

Development of an HSC-specific reporter line

The expression of the myelodysplastic syndrome 1 (*Mds1*) gene is highly enriched in LT-HSCs¹⁰. *Mds1* is transcribed from its own promoter in the *Mecom* locus, which also produces the well-known EVI1 gene product and the MDS1–EVI1 gene fusion product¹¹. We targeted an EGFP expression cassette to the first transcriptional start site of *Mds1* (Extended Data Fig. 1a). The resulting allele is predicted to be a hypomorph for MDS1 and MDS1–EVI1 but to have no effect on the expression of EVI1. Mice heterozygous for the GFP-linked allele (*Mds1*^{GFP/+}) showed normal haematopoietic parameters, frequency of HSCs and cell cycle properties, and response to myelosuppression (Extended Data Fig. 1b–f). Flow cytometric characterization of these

¹Stem Cell Program, Boston Children's Hospital, Boston, MA, USA. ²Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. ³Advanced Microscopy Program, Center for Systems Biology, Massachusetts General Hospital, Boston, MA, USA. ⁴Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, MA, USA. ⁵Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. ⁶Department of Bioengineering, University of California Merced, Merced, CA, USA. ⁷Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. ⁸Department of Molecular Biotechnology and Health Sciences, University of Torino, Torino, Italy. ⁹Dana Farber/Boston Children's Cancer and Blood Disorders Center, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. ¹⁰Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA. ¹¹Department of Chemistry, University of Pennsylvania, Philadelphia, PA, USA. ¹²Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Rochester, NY, USA. ¹³Present address: Novartis Institutes for BioMedical Research, Cambridge, MA, USA. ¹⁴These authors contributed equally: C. Christodoulou, J. A. Spencer, S. C. A. Yeh. *e-mail: charles_lin@hms.harvard.edu; fernando.camargo@childrens.harvard.edu

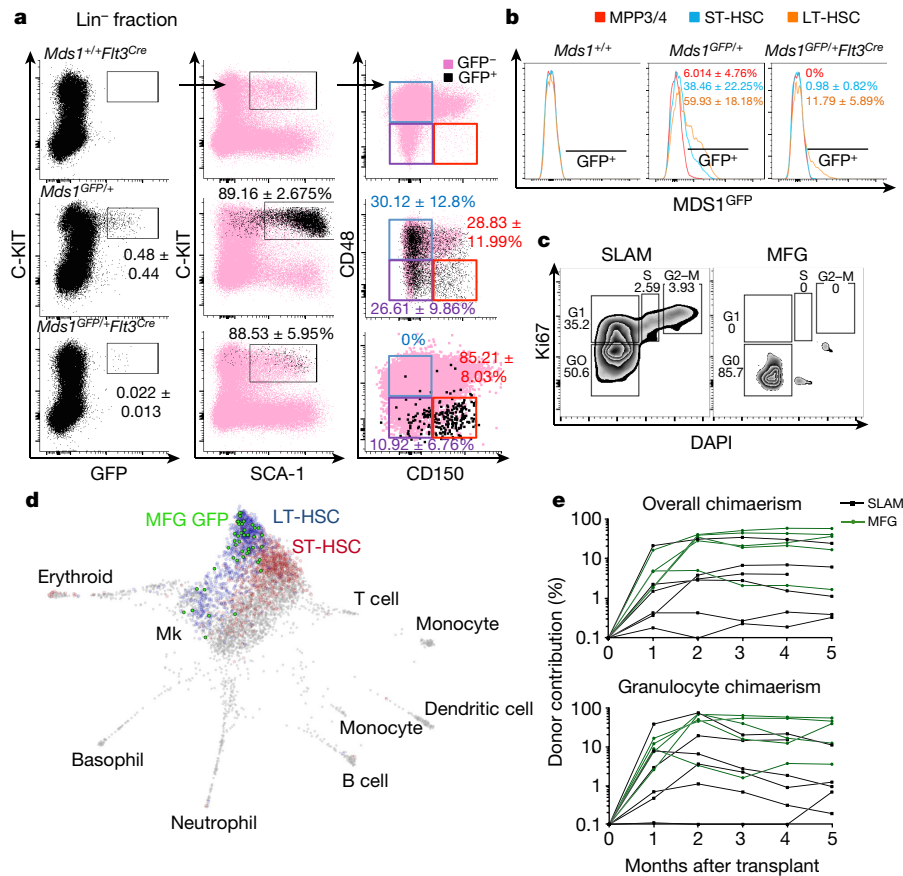


Fig. 1 | Generation and characterization of *Mds1*^{GFP/+} *Flt3*^{Cre} (MFG) mice. **a, b**, Flow cytometric analysis of *Mds1*^{GFP/+} only mice ($n = 10$) and *Mds1*^{GFP/+} *Flt3*^{Cre} mice ($n = 13$); mean \pm s.d. **c**, Cell cycle analysis of GFP⁺ cells from MFG mice versus HSCs isolated by LIN⁻ SCA1⁺ C-KIT⁺ CD150⁺ CD48⁺ (SLAM) immunophenotype. Representative analysis shown, depicting data from multiple mice (MFG, $n = 7$; SLAM, $n = 2$) that were pooled together to acquire the displayed data. **d**, SPRING plot layout of transcriptomes of 50 single MFG⁺ HSCs

projected in published scRNA dataset of HSCs and MPPs¹⁸. Blue, LT-HSCs ($n = 789$); red, ST-HSCs ($n = 742$); grey, other cells; bright green, MFG HSCs ($n = 46$). Each dot represents one cell. **e**, Overall and granulocyte chimaerism after transplantation in primary lethally irradiated recipients transplanted with 25 MFG⁺ ($n = 6$ mice) or SLAM cells ($n = 5$ mice) from *Mds1*^{GFP/+} *Flt3*^{Cre} mice. Each line represents an individual mouse. Only engrafted mice are shown.

mice confirmed the complete absence of GFP expression in any mature lineage-positive (LIN⁺) haematopoietic cells (Extended Data Fig. 2a, Supplementary File 2). GFP expression was predominantly restricted to a small fraction of CKIT⁺ SCA1⁺ cells (Fig. 1a). Using standard phenotypic parameters^{12,13}, we found that $28.83 \pm 11.99\%$ (mean \pm s.d.) of bone marrow cells gated solely on GFP could be categorized as LT-HSCs, $26.61 \pm 9.86\%$ as short-term HSCs (ST-HSCs) and $30.12 \pm 12.8\%$ as MPPs (Fig. 1a, Supplementary File 1). Within the phenotypic LT-HSC compartment, approximately 60% of cells expressed GFP, compared to 6% of MPPs (Fig. 1b). We refer to this mixed population as haematopoietic stem and progenitor cells (HSPCs). Notably, GFP was not expressed in non-haematopoietic compartments of the bone marrow (Extended Data Fig. 2b, c, Supplementary File 3).

With the aim of eliminating the labelling of MPPs in the *Mds1*^{GFP/+} model, we reasoned that the additional expression of a gene associated with early differentiation could facilitate exclusive identification of LT-HSCs. Increased brightness of the reporter in phenotypic LT-HSCs was inversely correlated with the expression of *Flt3*, a gene whose expression has been associated with the loss of long-term self-renewal^{14,15} (Extended Data Fig. 2d). Taking advantage of the fact that the GFP coding sequences in the *Mds1*^{GFP/+} allele are flanked by loxP sites (Extended Data Fig. 1a), we introduced a *Flt3*^{Cre} allele into our model (Extended Data Fig. 3a). This allele drives Cre-mediated recombination in cells, beginning in the ST-HSC compartment^{14,15} (Extended Data

Fig. 3b). Characterization of *Mds1*^{GFP/+} *Flt3*^{Cre} mice revealed an extremely rare GFP⁺ population (referred to as MFG cells) that corresponds to only $0.022 \pm 0.013\%$ of the lineage-negative bone marrow (Fig. 1a, b). Remarkably, approximately 85% of cells gated solely on the basis of GFP resided in the phenotypically defined LT-HSC fraction (Fig. 1a, Extended Data Fig. 3e). Another 10% of MFG GFP⁺ cells displayed slightly lower levels of CD150 and might be classified as ST-HSCs (Fig. 1a), and the other 5% represents CD150⁺ CD48⁺ cells that express lower levels of SCA1 and are likely to be megakaryocyte progenitors (MkPs; Extended Data Fig. 3c, Supplementary File 1). MFG cells constituted only about 12% of the phenotypic LT-HSC population (Fig. 1b). The specificity of LT-HSC labelling in MFG mice was recapitulated in bone marrow from multiple locations (Extended Data Fig. 3d). MFG cells represented a largely quiescent population (Fig. 1c, Extended Data Fig. 3g) that express relatively high levels of SCA1 and EPCR and little or no CD34 (Extended Data Fig. 3e, f), consistent with previously described dormant HSCs¹⁶.

To further validate our combined *Mds1*^{GFP/+} *Flt3*^{Cre} model, we performed single-cell RNA sequencing (scRNA-seq) in cells isolated exclusively on the basis of GFP expression¹⁷. The resulting transcriptomes were then extrapolated to a published single-cell transcriptional map of LT-HSCs, ST-HSCs, and multiple MPP populations (MPP2/3/4)¹⁸. Strikingly, virtually all MFG⁺ transcriptomes mapped to the most unprimed cluster of cells, in which phenotypic LT-HSCs also reside (Fig. 1d, Extended Data Fig. 4a). A small fraction of MFG⁺

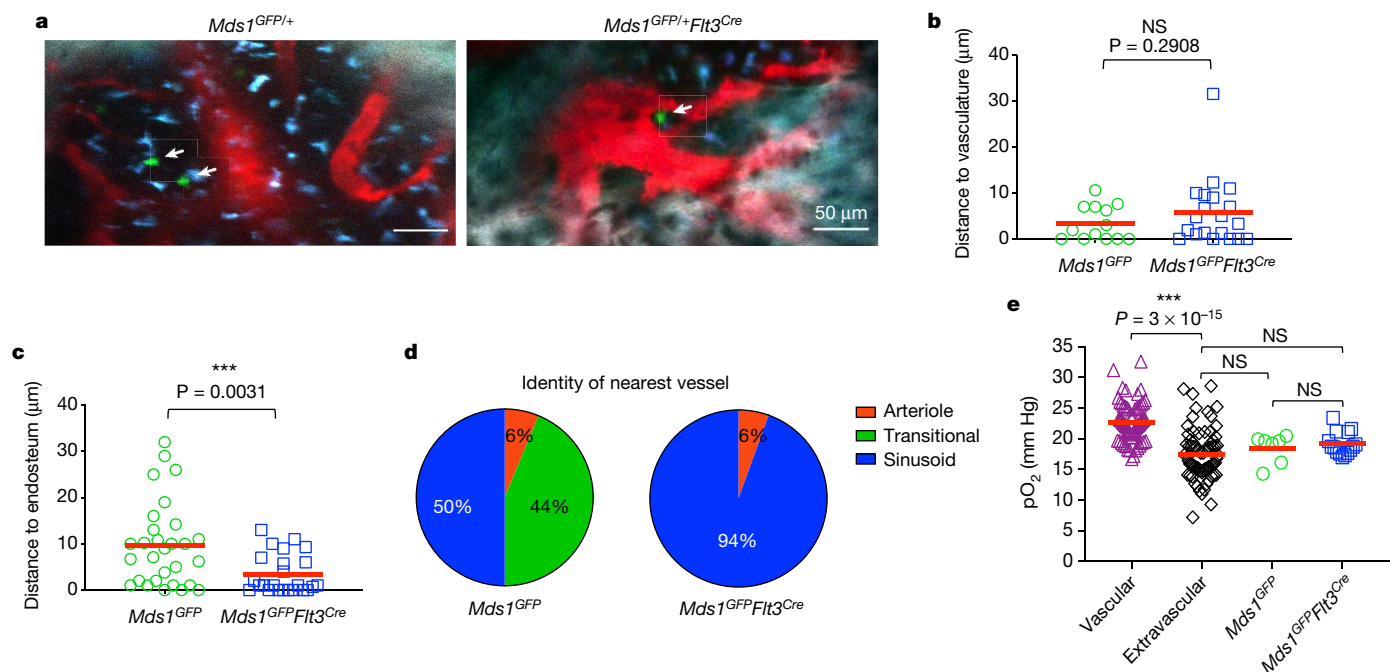


Fig. 2 | Steady-state localization and oxygen levels around MFG-HSCs and HSPCs. a, Representative intravital images of HSPCs (left, $n = 8$ mice) and an MFG-HSC (right, $n = 10$ mice) in the calvaria of *Mds1^{GFP/+}* and *Mds1^{GFP/+}Flt3^{Cre}* mice, respectively. GFP cells (white arrows) are shown in green, vasculature (Angiosense 680EX) in red, auto-fluorescence in blue, and bone (second harmonic generation) in white. Scale bars, $\sim 50 \mu\text{m}$. **b, c**, Distance from each HSPC ($n = 13$ and 29 cells from 3 and 4 mice for **b** and **c**, respectively) and

MFG-HSC ($n = 20$ and 24 cells from 6 and 8 mice for **b** and **c**, respectively) to the nearest vessel and endosteal surface, respectively, are displayed. **d**, Identity of nearest vessel for each HSPC ($n = 16$ cells) and MFG-HSC ($n = 18$ cells). **e**, Graph of in vivo oxygen measurements around individual HSPCs ($n = 2$ mice, 7 cells) and MFG-HSCs ($n = 2$ mice, 15 cells). P values calculated using two-tailed unpaired t -tests; red bars, mean.

cells showed megakaryocyte lineage priming (Fig. 1d, Extended Data Fig. 3c), which has recently been described in multipotent HSCs^{18,19}. This analysis also highlights the efficiency of our approach in restricting GFP expression to *Mds1⁺Flt3⁺* cells (Extended Data Fig. 4b). MFG cells expressed transcripts that were also enriched in dormant HSCs¹⁶ (Extended Data Fig. 4c, d). In addition, single-cell quantitative PCR analysis of a 280-gene haematopoietic gene panel²⁰ demonstrated clustering of MFG cells with LT-HSCs but no other progenitor signatures (Extended Data Fig. 4e). Finally, we performed long-term reconstitution assays to assess the potency of MFG cells in comparison to cells isolated using traditional flow cytometry markers for HSCs (LIN⁺SCA-1⁺C-KIT⁺CD150⁺CD48⁺, here referred to as SLAM cells) parameters. Limiting dilution transplants using 3–25 cells suggested that MFG-HSCs are at least as enriched as SLAM cells in transplantation capacity (Fig. 1e, Extended Data Fig. 4f). MFG cells also repopulated secondary recipients (Extended Data Fig. 4g). In addition, within the LIN⁺SCA-1⁺C-KIT⁺CD150⁺CD48⁺ compartment, long-term repopulating activity was enriched in cells expressing GFP (Extended Data Fig. 4h). Thus, our MFG animal model allows the isolation of a highly quiescent sub-population of LT-HSCs with potent repopulation potential.

Localization of MFG-HSCs in the calvaria

Using these two reporter models, we performed imaging of GFP⁺ cells in the calvaria of live mice^{3,4}. As expected, MDS1-GFP HSPCs were more prevalent than MFG-HSCs (Fig. 1a, b, 2a). Both cell types were located peri-vascularly at an average distance of less than $10 \mu\text{m}$ from the closest vessel (Fig. 2b, Supplementary Videos 1, 2). MFG-HSCs were also found similarly close to the endosteum (Fig. 2c), pointing to a possible dual endosteal-vascular niche, as suggested previously^{3,21,22}. However, we found that MFG-HSCs were almost exclusively associated with sinusoids rather than arterioles (Fig. 2d). While HSPCs also predominantly

localized close to sinusoids, a significant fraction of these were also near transition zone vessels (Fig. 2d) and their distance from the endosteum was more varied (Fig. 2c), suggesting that MFG-HSCs and downstream HSPCs occupy different micro-niches.

Given the known developmental and structural differences between flat and long bones²³, we also imaged femurs using a quantitative deep-imaging protocol²⁴. We identified a very small number of GFP⁺c-Kit⁺ HSCs in 250- μm -thick, whole-bone femoral sections from MFG mice (Extended Data Fig. 5a–e). Approximately 70% of MFG-HSCs were located within $5 \mu\text{m}$ of sinusoidal CD105⁺ cells, but this was not statistically significant in comparison to random dots (Extended Data Fig. 5f). In addition, MFG-HSCs did not differ significantly from random spots in their distance to the endosteum (about 12% were within $10 \mu\text{m}$ and more than 50% were over $50 \mu\text{m}$ away; Extended Data Fig. 5b, d, f), underscoring the difference between the calvarium and the long bone, particularly the diaphysis.

MFG-HSCs are not found in deep hypoxic zones

Low oxygen tension (hypoxia) has been historically thought to be a shared niche characteristic that is critical for maintaining stem cell quiescence²⁵. However, support for the existence of a hypoxic niche has largely come from indirect evidence and measurements lacking spatial resolution²⁶. Using an oxygen sensor and two-photon phosphorescence lifetime microscopy²⁷, we measured the local pO₂ surrounding individual HSPCs and MFG-HSCs in their native microenvironments (Extended Data Fig. 6a–f). First, we confirmed the overall hypoxic status of the calvarial bone marrow²⁷, with intravascular pO₂ in the range of 15–30 mm Hg (mean ~ 23 mm Hg, about 3% O₂) and extravascular pO₂ in the range of 10–25 mm Hg (mean ~ 17 mm Hg, about 2% O₂; Fig. 2e). We then measured pO₂ around individual HSPCs and MFG-HSCs, and found similar oxygen levels (~ 18 and ~ 19 mm Hg) close to the average

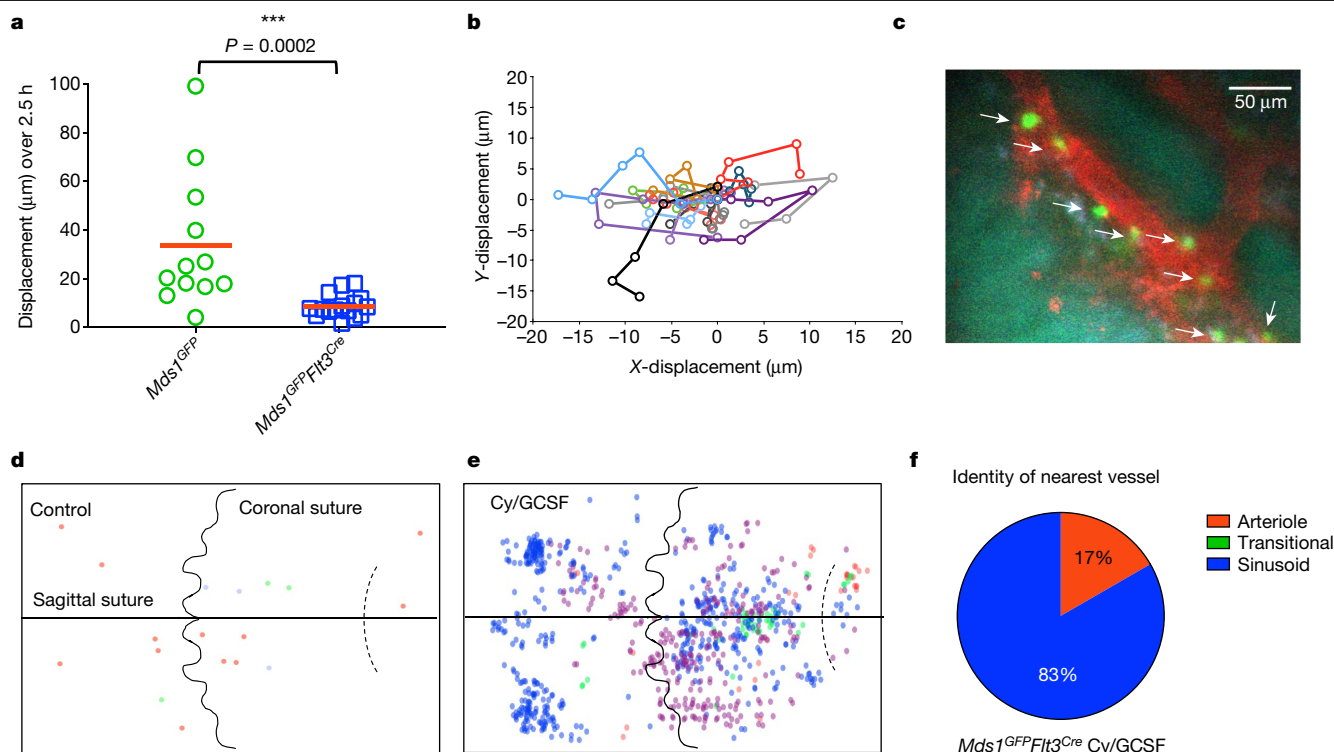


Fig. 3 | Increased motility, expansion, and localization of activated MFG-HSCs. **a**, In vivo motility measurements of HSPCs ($n = 12$ cells) and MFG-HSCs ($n = 16$ cells) at steady-state over a 2.5-h imaging period. Red bars, mean. P value calculated using two-tailed Mann–Whitney test. **b**, Cell tracks for 16 MFG-HSCs over a 2.5-h imaging period. Images were acquired every 30 min. **c**, Representative intravital image of a Cy/G-CSF-treated MFG mouse 4.5 days after the beginning of treatment. MFG cells (green), vasculature (red,

Angiosense 680EX), auto-fluorescence (blue). Arrows, GFP⁺ cells. The experiment was performed four times with similar results. **d**, **e**, Graphical map of the locations of MFG-HSCs in the calvaria of untreated and Cy/G-CSF-treated mice ($n = 3$ and 4, respectively). Location data from individual mice are indicated by different colours. **f**, Identity of nearest vessel for each MFG⁺ cell ($n = 12$ cells) after treatment with Cy/G-CSF. Compare to untreated mice in Fig. 2d.

extravascular pO₂ in the bone marrow (Fig. 2e). Thus, although we detected regions in the bone marrow with pO₂ as low as 10 mm Hg, HSPCs and MFG-HSCs were not found in these regions, suggesting that localization to the regions of deepest hypoxia is not a prerequisite for maintenance of MFG-HSC quiescence.

Heterogeneous HSC response to activation

We next examined the dynamic behaviours of HSPCs and LT-HSCs in their native niches. In vivo time-lapse imaging of the calvarium revealed that MFG-HSCs displayed low baseline motility whereas HSPCs showed enhanced motility (Fig. 3a, b, Supplementary Videos 3–6). To assess whether HSC behaviours would be affected in the context of activation, we used a cyclophosphamide (Cy)/G-CSF protocol that leads to expansion and subsequent mobilization of LT-HSCs²⁸ (Extended Data Fig. 7a). Fluorescence-activated cell sorting (FACS) and imaging analysis demonstrated a tenfold increase in the number of MFG cells after treatment (Fig. 3c–e, Supplementary Video 7, Extended Data Fig. 7b, c). MFG cells are still enriched in the phenotypic LT-HSC fraction in this activated state (Extended Data Fig. 7b) and display a concomitant increase in cell cycle activity (Extended Data Fig. 7d). Time-lapse imaging of treated animals for about 6 h after the third dose of G-CSF (Fig. 3c, Supplementary Video 8) showed that, on average, MFG-HSCs became significantly more motile ($P = 0.0001$), with displacement measurements even higher than those of steady-state HSPCs (Extended Data Fig. 7e, Supplementary Video 8), but the response was heterogeneous, ranging from cells displaying limited displacement to a small number of cells that fully exited the bone marrow into the blood stream (Supplementary Video 8, Extended Data Table 1). Similarly, following treatment with the myeloablative agent

5-fluorouracil (5-FU), the number of MFG⁺ cells increased (Extended Data Fig. 8a, Supplementary Videos 9, 11) and a subset exhibited higher motility, particularly on day 20 after treatment (Extended Data Fig. 8b, Supplementary Videos 10, 12). These data suggest that enhanced motility is a common feature of the HSC response to injury, although we cannot rule out the possibility that the response is a result of indirect action on the niche by Cy/G-CSF or 5-FU.

Activated MFG-HSCs were found, on average, to be further away from the endosteum than native MFG-HSCs (Extended Data Fig. 7f). Notably, they were even closer to the vasculature, with an average distance of about 1 μ m (Extended Data Fig. 7g), and maintained their sinusoidal proximity (Fig. 3f). By assessing the distribution of activated MFG-HSCs in the entire calvarial region, we identified unique patterns of HSC proliferation. First, native MFG-HSCs were found as rare single cells within the bone marrow, whereas activated MFG-HSCs appeared as clusters (Fig. 3c–e), suggesting that MFG-HSC proliferation occurs within spatially restricted domains. Second, a subset of MFG⁺ cells remained as single cells while others formed clusters in both the Cy/G-CSF and 5-FU models (Fig. 3e, Extended Data Fig. 8c), suggesting that the proliferative response is heterogeneous among HSCs. To assess whether these clusters were clonal, we generated *Mds1^{CreER/+}Rosa26^{Confetti/+}* mice²⁹ (Extended Data Fig. 8d–f). In untreated mice, labelled cells were usually found as rare single cells of different colours dispersed throughout the bone marrow (Extended Data Fig. 8g). Following treatment with Cy/G-CSF, we observed labelled cell clusters made up predominantly of cells of a single colour (Extended Data Fig. 8g, h). Quantitative analysis confirmed that labelled nearby cells were more likely to have the same colour than a mixture of colours (Extended Data Fig. 8h), providing evidence for clonal HSPC proliferation within confined physical domains.

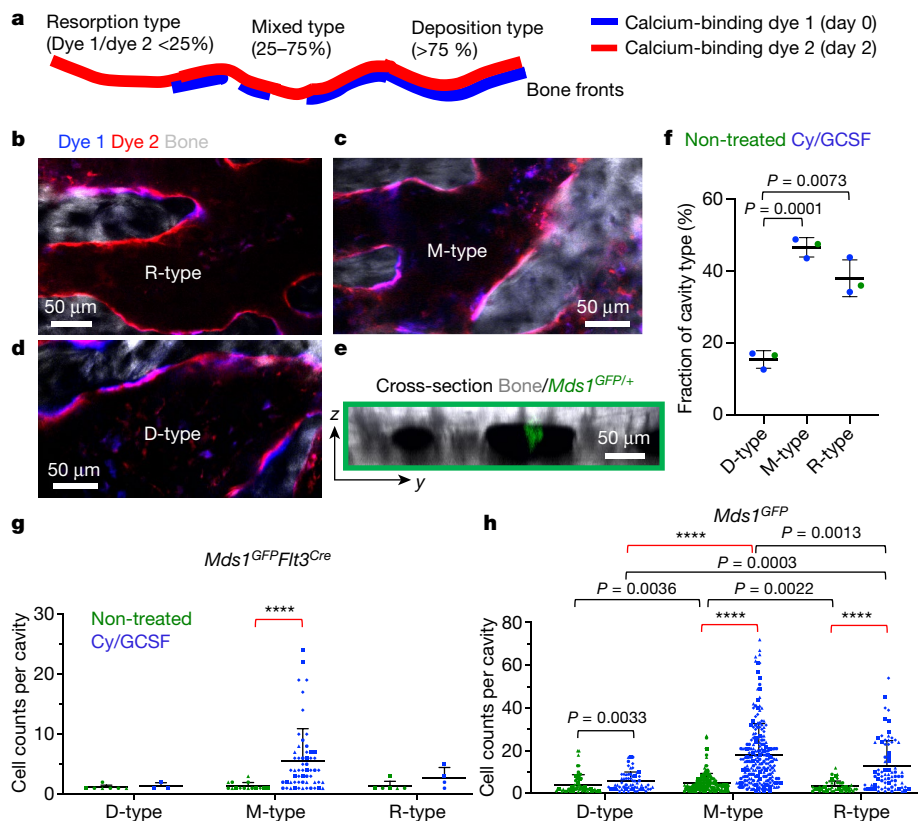


Fig. 4 | Heterogeneity of bone remodelling stages governs expansion of MFG-HSCs (*Mds1*^{GFP/Flt3^{Cre} mice) and HSPCs (*Mds1*^{GFP/+} mice). **a, The double calcium staining strategy that identifies D-, M- and R-type cavities. Dye 1, delivered 48 h before imaging, shows the old bone front that has been eroded to varying extents; dye 2, delivered before imaging, shows the new bone front. **b–d**, Expanded views, showing distinct cavity types defined by the dye 1:dye 2 pixel ratios. **e**, A sagittal section of bone marrow cavities containing *Mds1*^{GFP/+} cells. **f**, Fractions of D-, M- and R-type cavities in the calvaria of non-treated or}**

treated mice (two-tailed *t*-test, $n = 155$ cavities from non-treated and 80 or 73 bone marrow cavities from treated animals ($n = 3$ mice); mean \pm s.d.).

g, Quantification of MFG-HSCs in D-, M- or R-type cavities at steady-state and after Cy/GCSF activation. $n = 4$ mice per group, plotted as different symbols; black line represents mean \pm s.d. **h**, Quantification of HSPCs in D-, M- or R-type cavities at steady-state and after Cy/GCSF activation. $n = 4$ mice per group, plotted as different symbols. Two-sided Mann–Whitney test used unless otherwise specified, **** $P < 0.0001$, black line represents mean \pm s.d.

HSC expansion restricted by bone remodelling

The observation of heterogeneous HSC proliferation in restricted physical domains prompted us to re-examine the characteristics of the microenvironment that either support clonal expansion or maintain cell quiescence. Recognizing that the bone is constantly undergoing remodelling, we hypothesized that the stages of bone turnover impose an additional degree of heterogeneity in the bone marrow microenvironment that is not captured by the prevailing view centred on the endosteal versus perivascular duality. To visualize the stages of bone turnover, we administered two (spectrally distinct) calcium-binding dyes³⁰ 48 h apart, and imaged the calvarium immediately after the second dye injection. The two dyes mark the positions of the old and new bone fronts, respectively, and reveal where the old bone front has been eroded (Fig. 4a). We quantified the ratio of the two dyes and classified the cavities as D-type (undergoing predominantly bone deposition), R-type (predominantly bone resorption), and M-type (mixed). We confirmed that osteoblasts are biased towards D-type cavities while osteoclasts are biased towards R-type cavities. A mixture of osteoblasts and osteoclasts were found at intermediate levels in M-type cavities (Extended Data Fig. 9a–g). Using this double-staining scheme (Fig. 4b–e), we quantified the fractions of D-, M-, R- cavities in the calvaria (Fig. 4f), as well as the spatial distributions of native MFG-HSCs and HSPCs. During the steady state, MFG-HSCs were found in baseline numbers in all cavity types, while HSPCs tended to be enriched in M-type

cavities (Fig. 4g, h). After activation with Cy/GCSF, however, expanded MFG cells were found almost exclusively in a subset of M-type cavities (Fig. 4g, Supplementary Video 13, Extended Data Fig. 10a). HSPCs were also found to expand preferentially in M-type cavities after activation, although this preference was less pronounced (Fig. 4h, Supplementary Video 14, Extended Data Fig. 10b). This evidence of heterogeneity in types of bone marrow cavity, and of a subset of M-type cavities that favours HSC expansion, supports our earlier observation that HSCs expand clonally in restricted physical domains.

Discussion

Our work here describes the generation and characterization of an animal model in which a single-colour reporter can be used for the identification and live imaging of LT-HSCs in the native niche without transplantation (Extended Data Table 2). We found evidence of heterogeneity in both the HSC response to injury and the bone marrow microenvironment, coupled to the stages of bone remodelling, that has not been recognized previously to our knowledge^{3,27,31}. Notably, we also found distinct cavity types in the metaphyses of long bones (Extended Data Fig. 11 a–f, Supplementary Video 15). The existence of distinct types of bone marrow cavity implies that the traditional way of characterizing the HSC niche as endosteal or perivascular is inadequate, as the microenvironment, including the perivascular niches, contained within these cavities is likely to differ depending on the local calcium

gradient³² and the downstream effects of osteoclast degradation³³. To fully characterize the regulatory factors that govern HSC quiescence versus proliferation, it will be necessary to develop molecular profiling technology³⁴ that can spatially map distinct bone cavities.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1971-z>.

1. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
2. Busch, K. & Rodewald, H. R. Unperturbed vs. post-transplantation haematopoiesis: both *in vivo* but different. *Curr. Opin. Hematol.* **23**, 295–303 (2016).
3. Lo Celso, C. et al. Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457**, 92–96 (2009).
4. Lo Celso, C., Lin, C. P. & Scadden, D. T. In vivo imaging of transplanted haematopoietic stem and progenitor cells in mouse calvarium bone marrow. *Nat. Protoc.* **6**, 1–14 (2011).
5. Sipkins, D. A. et al. In vivo imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435**, 969–973 (2005).
6. Cao, X. et al. Irradiation induces bone injury by damaging bone marrow microenvironment for stem cells. *Proc. Natl Acad. Sci. USA* **108**, 1609–1614 (2011).
7. Acar, M. et al. Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature* **526**, 126–130 (2015).
8. Chen, J. Y. et al. Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. *Nature* **530**, 223–227 (2016).
9. Gazit, R. et al. Fgd5 identifies haematopoietic stem cells in the murine bone marrow. *J. Exp. Med.* **211**, 1315–1331 (2014).
10. Zhang, Y. et al. PR-domain-containing Mds1-Evi1 is critical for long-term haematopoietic stem cell function. *Blood* **118**, 3853–3861 (2011).
11. Métais, J. Y. & Dunbar, C. E. The MDS1-EVI1 gene complex as a retrovirus integration site: impact on behavior of haematopoietic cells and implications for gene therapy. *Mol. Ther.* **16**, 439–449 (2008).
12. Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of haematopoietic stem cells and multipotent progenitors. *Cell Stem Cell* **13**, 102–116 (2013).
13. Pietras, E. M. et al. Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell* **17**, 35–46 (2015).
14. Boyer, S. W., Schroeder, A. V., Smith-Berdan, S. & Forsberg, E. C. All haematopoietic cells develop from haematopoietic stem cells through Flk2/Flt3-positive progenitor cells. *Cell Stem Cell* **9**, 64–73 (2011).
15. Buza-Vidas, N. et al. FLT3 expression initiates in fully multipotent mouse haematopoietic progenitor cells. *Blood* **118**, 1544–1548 (2011).
16. Cabezas-Wallscheid, N. et al. Vitamin A-retinoic acid signaling regulates haematopoietic stem cell dormancy. *Cell* **169**, 807–823.e819 (2017).
17. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
18. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
19. Sanjuan-Pla, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
20. Guo, G. et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* **13**, 492–505 (2013).
21. Kunisaki, Y. et al. Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* **502**, 637–643 (2013).
22. Nombela-Arrieta, C. et al. Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. *Nat. Cell Biol.* **15**, 533–543 (2013).
23. Lassailly, F., Foster, K., Lopez-Onieva, L., Currie, E. & Bonnet, D. Multimodal imaging reveals structural and functional heterogeneity in different bone marrow compartments: functional implications on haematopoietic stem cells. *Blood* **122**, 1730–1740 (2013).
24. Coutu, D. L., Kokkalis, K. D., Kunz, L. & Schroeder, T. Multicolor quantitative confocal imaging cytometry. *Nat. Methods* **15**, 39–46 (2018).
25. Takubo, K. & Suda, T. Roles of the hypoxia response system in haematopoietic and leukemic stem cells. *Int. J. Hematol.* **95**, 478–483 (2012).
26. Parmar, K., Mauch, P., Vergilio, J. A., Sackstein, R. & Down, J. D. Distribution of haematopoietic stem cells in the bone marrow according to regional hypoxia. *Proc. Natl Acad. Sci. USA* **104**, 5431–5436 (2007).
27. Spencer, J. A. et al. Direct measurement of local oxygen concentration in the bone marrow of live animals. *Nature* **508**, 269–273 (2014).
28. Morrison, S. J., Wright, D. E. & Weissman, I. L. Cyclophosphamide/granulocyte colony-stimulating factor induces haematopoietic stem cells to proliferate prior to mobilization. *Proc. Natl Acad. Sci. USA* **94**, 1908–1913 (1997).
29. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
30. Yeh, S. A., Wilk, K., Lin, C. P. & Intini, G. In vivo 3D histomorphometry quantifies bone apposition and skeletal progenitor cell differentiation. *Sci. Rep.* **8**, 5580 (2018).
31. Rashidi, N. M. et al. In vivo time-lapse imaging shows diverse niche engagement by quiescent and naturally activated haematopoietic stem cells. *Blood* **124**, 79–83 (2014).
32. Adams, G. B. et al. Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature* **439**, 599–603 (2006).
33. Kollet, O. et al. Osteoclasts degrade endosteal components and promote mobilization of haematopoietic progenitor cells. *Nat. Med.* **12**, 657–664 (2006).
34. Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Mice and genotyping

The generation of *Mds1*^{GFP/+} mice was generated by cloning and homologous recombination of the linearized targeting vector via electroporation into v6.5 embryonic stem (ES) cells. After selection with neomycin and clonal screening by PCR, correctly targeted ES cell clones were injected into C57Bl/6 blastocysts. Derived chimaeras were initially bred to C57Bl/6 to obtain germline transmission, followed by crossing with FLPe mice³⁵ to remove the Frt-Neo-Frt cassette that was part of the original targeting vector. Derived mice were backcrossed onto a C57Bl/6 background for more than six generations and mice were analysed via PCR to identify their genotype ('5'-AGAGTGAAGACCGAGTGTGTG-3', '5'-GTACAGGGTAGGCTGCTCAACT-3', '5'-CTCCCTCCCAGCTTTTGTCT-3'). Some of the displayed data comes from mice that still carried the Frt-Neo-Frt cassette, which showed slightly lower mean fluorescence intensity in bone marrow cells. A similar strategy was used to generate the *Mds1*^{CreER} allele. For all experiments, 2–12-month-old adult mice of both sexes were used and wild-type littermates were used as controls. *Flt3*^{Cre} mice^{14,15}, Rosa26-CAG-loxp-stop-loxp-tdTomato reporter mice³⁶ and Rosa26-CAG-loxp-stop-loxp-Confetti reporter mice²⁹ have been described. For identification of *Flt3*^{Cre} and *Mds1*^{CreER}, Cre primers were used ('5'-TTACTGACCGTACACAAAATTTGCC-3', '5'-CTGGCAGCGATCGCTATTTCCATG-3'). Mice were bred and housed according to NIH guidelines in our AAALAC-accredited, specific-pathogen-free animal care facilities at Boston Children's Hospital or Massachusetts General Hospital. 2.3Collagen1α1-GFP (2.3Col1-GFP) mice³⁷ were generously provided by Dr. Jayaraj Rajagopal (Massachusetts General Hospital). All animal protocols were approved by the Animal Resources at Children's Hospital Boston, Boston Children's Hospital Institutional Animal Care and Use Committee, and Massachusetts General Hospital Institutional Animal Care and Use Committee. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All results involved in the study were acquired according to ethical standards.

HSC isolation, flow cytometry and cell sorting

Bone marrow cells were isolated by crushing of the bones using a mortar and pestle in Ca²⁺/Mg²⁺-free phosphate-buffered saline (D-PBS) supplemented with 2% fetal bovine serum (FBS) and 1× penicillin/streptomycin (Pen/Strep) (Invitrogen). Viable cell number was calculated by manually counting with a haemocytometer or using a TC20 Automated Cell Counter (Bio-Rad). The cell suspension was filtered through a 70-μm strainer. For HSC identification via flow cytometry, the cells were stained for C-KIT (eBioscience), SCA-1 (eBioscience, BioLegend), CD48 (BD Pharmingen) and CD150 (BioLegend) as well as a lineage marker cocktail consisting of B220, TER-119, GR-1, CD4 and CD8α (eBioscience). For experiments requiring lineage depletion, antibody staining for B220, Ter119, GR-1, CD4 and CD8α biotin-conjugated antibodies was first performed followed by application of anti-biotin beads (Miltenyi Biotec) and depletion using magnetic separation columns (Miltenyi Biotec). For megakaryocyte progenitor staining, the cells were stained for lineage marker cocktail, C-KIT, SCA-1, CD150, CD41 (BioLegend) and FCyR (eBioscience). For identification of common myeloid progenitors (CMPs), granulocyte-monocyte progenitors (GMPs) and megakaryocyte-erythroid progenitors (MEPs), cells were stained for lineage marker cocktail, C-KIT, SCA-1, CD34 (eBioscience) and FCyR. For mesenchymal stem cells, cell suspension was stained for the lineage cocktail, CD45, PDGFRα and integrin-αV (eBioscience). For endothelial cells, lineage cocktail, CD45, CD31 and VE-cadherin (eBioscience) were used. For identification of pre and pro B cells, immature B cells and mature B cells, B220 and IgM (eBioscience) were used. For erythroid cells, the primary marker was Ter-119 (eBioscience); for monocytes and neutrophils, MAC-1 (eBioscience) and LY6-G (BD Pharmingen) were used; and for T cells CD4 and CD8 were

used. Antibody staining of cell suspensions was always performed on ice for 45 min. 4',6-diamidino-2-phenylindole (DAPI, 10 μg/ml in PBS; Invitrogen) was used for exclusion of dead cells during flow cytometry. Relevant flow cytometry gating strategies for the identification of different mature cell populations, LT-HSCs, ST-HSCs, MPP2s, MPP3/4s and endothelial cells are available in the Supplementary Information. Transplanted cells were double-sorted to increase purity. For transplantation of low cell numbers, all secondary sorts were performed in a plate using the automated plate reader sorting function. For FACS analysis, a BD LSR II Flow Cytometer was used, while cell sorting was performed using a BD FACS Aria II sorter (BD Biosciences). Flow cytometry data were analysed using FlowJo (Tree Star).

Cell cycle analysis

As each animal contained an average of only 600–700 MFG⁺ sorted cells, the cell cycle analysis demonstrated in Fig. 1c and Extended Data Fig. 3g represent GFP cells isolated from seven *Mds1*^{GFP/+}*Flt3*^{Cre} mice; thus, the data represent an average from seven mice that were pooled together. Upon identification and sorting purification of corresponding cellular populations as described above, cells were fixed in ice cold 70% ethanol. Cells were then washed and stained with Ki67 (BioLegend) for 30 min on ice to distinguish G0/G1 phase. DAPI was finally used for staining and analysis of G1/G2 versus M/S phase. Cell cycle analysis was performed using a BD FACS Aria II sorter.

Competitive reconstitution assays in irradiated mice and peripheral blood analysis

Bone marrow transplant recipients were 8–12-week-old B6.SJL-Ptprc^a Pepc^b/BoyJ (CD45.1) mice. Before transplantation, mice were lethally irradiated using a gamma irradiator with a split dose of 11 Gy with a 3-h interval between the two doses. Cells were transplanted via retro-orbital injection into anaesthetized mice. One hundred thousand whole bone marrow CD45.1 cells were used as competitors unless stated otherwise. For limiting dilution studies, HSC frequency was calculated using Extreme Limiting Dilution Analysis software (<http://bioinf.wehi.edu.au/software/elda/>)³⁸ with data taken 4 months after transplantation. The lower stem cell frequency reported here might be due to the incomplete backcrossing of MFG mice (six generations), the presence of constitutive CRE in haematopoietic cells and/or technical reasons. For secondary transplants, two million whole bone marrow cells from primary recipients were transplanted into lethally irradiated secondary recipients. For blood analysis of transplanted recipients, blood was collected at 4-week intervals for at least 16 weeks after transplantation. Peripheral blood was first treated with red blood cell lysis buffer to remove red blood cells followed by antibody staining for B cells (CD19, eBioscience), T cells (CD4, CD8α) and granulocytes (Ly6-G). The percentage chimaerism was estimated using CD45.1 (BioLegend) and CD45.2 (eBioscience) antibody staining.

Blood cell counts and treatment with 5-FU, Cy/G-CSF and tamoxifen

Blood samples were collected via the retro-orbital vein in EDTA-coated tubes. Blood cell counts were performed using a HEMAVET950 (Drew Scientific) cell blood counter. For blood cell kinetic analysis upon 5-FU treatment, cell counts were performed on day 0, 3, 7, 10, 13 and 17. 5-FU was delivered via retro-orbital injection as single dose of 150 mg/kg immediately after day 0 and a bleeding sample was collected while control mice were injected with PBS. In addition, bone marrow from treated mice was analysed using flow cytometry or imaging was performed at the indicated time points. For Cy/G-CSF experiments, cyclophosphamide was delivered via intraperitoneal injection as a single dose of 200 mg/kg on day 1, followed by subcutaneous injection of G-CSF on days 2, 3 and 4 at 250 μg/kg per day followed by bone marrow flow cytometry analysis or live animal imaging of the calvarial bone marrow. For *Mds1*^{CreER/+}*Rosa26*^{Confetti/+} mouse experiments, cyclophosphamide

or PBS was administered on day 1 followed by GCSF and tamoxifen injection on day 2 and further GCSF or PBS administration on days 3 and 4 according to each experiment. Tamoxifen was administered at a 2 mg dose via intraperitoneal injection to target labelling to the LT-HSC compartment. Bone marrow analysis of femurs and calvaria was performed 24 h after the final GCSF dose according to each experiment.

Single cell inDrops RNA sequencing

GFP cells were sorted from *Mds1^{GFP/+} Flt3^{Cre}* mice and single cells were encapsulated using droplet microfluidic technology as previously described¹⁷. Although 1,200 GFP⁺ cells were sorted, only about 400 cells were encapsulated; the loss of more than half of the population is standard for low cell number populations in the inDrop platform. Upon library preparation of barcoded single cells, RNA sequencing was performed. To process the data, we used a previously published workflow and code available at <https://github.com/AllonKleinLab/SPRING>. Ensembl release 81 mouse mm10 cDNA plus the sequence of loxP-IRES-GFP-polyA-loxP was used as reference. SPRING plots were generated using the next four-step process. Initially, cells with few mRNA counts (<1,000 unique molecular identifiers) and stressed cells (mitochondrial gene-set Z-score >1) were filtered out. The remaining high-quality cells were total-counts normalized. We next filtered genes, keeping those that were well detected (mean expression >0.05) and highly variable (CV >2). Finally, the data were normalized by Z-scoring each gene and applying principal components analysis (PCA), retaining the top 50 PCs. Following filtration and bioinformatics analysis, only 50 GFP⁺ cells passed quality control and were used for plotting. The data acquired from the GFP⁺ cells were then plotted with previously published data for LSK cells¹⁸ upon transformation in the PCA space of the previously published data. In brief, the two datasets were integrated using the library sklearn.decomposition.PCA (python 2.7). The fit function was used to calculate the first 50 PCs for the single cell LSK dataset¹⁸. Then, the normalized and filtered count matrices of the GFP and LSK cells were vertically combined. Before combining the two matrices, the GFP matrix was scaled in order to have a comparable amount of normalized counts in correlation to the LSK matrix. The resulting Z-score-combined matrix was used as input for the transform function to project the combined dataset onto the original LSK dataset. The output generated by the transform function and the corresponding distance matrix, which was obtained using the SPRING function `get_distance_matrix`, were used to generate the final SPRING plot. All data were visualized as previously described¹⁸. This reduced any batch effects between the two experiments. The coordinates generated by the SPRING plots were plotted using R.

Single cell fluidigm analysis

GFP cells were primarily sorted from *Mds1^{GFP/+} Flt3^{Cre}* mice followed by secondary single cell sorting directly in 96-well plates containing PCR buffer. Sorted plates were frozen on dry ice followed by reverse transcription, pre-amplification and high-throughput microfluidic real-time PCR for 180 transcription factors as previously described²⁰. Data analysis and hierarchical clustering were performed using MultiExperiment Viewer (MeV) program. Previously published data for CMPs, GMPs, MEPs, common lymphoid progenitors (CLPs), MPPs and LT-HSCs²⁰ using the same 180 real-time PCR platform were overlaid for comparison to the GFP cells.

Synthesis and characterization of phosphorescent Oxyphor PtG4 probe

The structure of Oxyphor PtG4 is almost identical to that of the previously published Oxyphor PdG4 probe³⁹. The synthesis of the core porphyrin and the synthesis of a dendritic probe similar to Oxyphor PtG4 have been published previously⁴⁰. All synthesis steps were identical to those developed for the synthesis of PdG4⁴⁰. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry

was used to confirm the identity of the intermediate products and of the target probe molecule. Calibration, assessment of phosphorescence oxygen quenching and absorption spectra measurements for the Oxyphor PtG4 probe were performed as previously described^{39,41}.

In vivo and ex vivo imaging

All in vivo imaging experiments were performed according to procedures approved by the Institutional Animal Care and Use Committee at Massachusetts General Hospital. In brief, mice were either anaesthetized with an induction dose of 3–4% isoflurane (96% O₂) and a maintenance dose of 1.25–2% isoflurane or given an intraperitoneal bolus injection of ketamine (100 mg/kg) and xylazine (15 mg/kg). Animals were deemed anaesthetized by the toe pinch method. To minimize pain, mice were treated with buprenorphine (0.05–0.1 mg/kg). The hair on the calvarium was removed with scissors or a mechanical trimmer and then the skin was wiped with alcohol. Next, a calvarial skin flap was created with a U-shaped incision to reveal the underlying calvarial bone as previously described²⁷. Mice were injected with imaging agents (for example, vascular labels) retro-orbitally, mounted in a custom-designed heated mouse holder, and secured to the stage of a home-built multiphoton/confocal laser-scanning video-rate microscope (for z-stack or time-lapse imaging) or an Olympus FVMPE-RS multiphoton imaging platform (for oxygenation measurements)²⁷. A drop of 0.9% saline was applied to the skull to act as the immersion fluid, and a Zeiss 63×1.15 numerical aperture water-dipping objective, an Olympus 60×1.0 numerical aperture water-dipping objective, or an Olympus 25×1.05 numerical aperture water-dipping objective was used for all imaging. For endpoint imaging, mice were killed while under anaesthesia using approved procedures. For survival imaging, the skin flap was closed with 6-0 vinyl sutures (Ethicon). Triple antibiotic ointment (bacitracin, neomycin, and polymyxin-B sulfate) was applied to the top of the surgical site to minimize the chance of infection. Mice were put in a heated cage and monitored until fully awake. For 6-h imaging sessions, mice were given an intraperitoneal injection of ~100 µl 0.90% saline solution every hour to ensure proper hydration.

GFP was excited at 491 nm (confocal) or 950 nm (two-photon) and collected at ~505–540 nm using a photomultiplier tube. Angiosense 680EX (~100 µl at 2 nmol/100 µl, Perkin-Elmer) for labelling the vasculature was excited at 635 nm (confocal only) and collected at ~665–725 nm using a photomultiplier tube. Autofluorescence generated from the 491 nm or 950 nm excitations was collected at ~570–610 nm using a photomultiplier tube. Second harmonic generation (SHG) from collagen in the bone was excited at 775 nm or 840 nm (two-photon only) and collected at ~340–460 nm with a photo-multiplier tube. Phosphorescence was excited at 1,150 nm by a Ti:Sa femtosecond laser (Insight, Spectra-Physics) and collected above 750 nm with a photomultiplier tube. For calcium staining of endosteal bone fronts, calcein blue (30 mg/kg), tetracycline (35 mg/kg, Sigma), and Alizarin red S (40 mg/kg), were excited at 775 nm and collected at 415–455 nm, 500–550 nm, and 580–650 nm, respectively. Rhodamine B dextran 70 kDa (0.5 mg/50 µl, Sigma) was used as a vascular contrast with CatK 680 FAST (2 nmol/100 µl injected 6 h before imaging, Perkin-Elmer) for labelling osteoclasts, excited simultaneously at 532 nm and 638 nm and collected at 570–620 nm and 665–745 nm.

For steady-state in vivo imaging, 15–60 frames from the live video mode were averaged to acquire single 500 × 500 pixel images. Z-stacks were acquired with 1–2-µm steps and time-lapse images were acquired at 30-s intervals for 20 min or longer. For Cy/GCSF in vivo imaging, z-stacks were acquired with 2-µm steps every 20 min for ~6 h. Calvarial cell location maps in Fig. 3d, e were created in Matlab using custom code based on the x, y, and z coordinates of each cell. Data from each mouse were aligned and then overlaid using the locations of the coronal and sagittal sutures.

For pO₂ measurements, ~75 µl of 1.7 mM Pt-G4 suspended in 0.9% PBS (1× PBS, Invitrogen) was injected intravenously before imaging.

Article

In each pO_2 measurement location, multiple pulse excitation/emission cycles were used to record the phosphorescence lifetime decay. For each cycle, the probe was excited for $\sim 10\text{--}20\ \mu\text{s}$ followed by $\sim 150\ \mu\text{s}$ for time-resolved photon collection. Quantitative pO_2 values were obtained by fitting the phosphorescence decay with a single exponential to get an average lifetime of phosphorescence. This lifetime value was converted to pO_2 using an in vitro calibration curve for the same batch of Pt-G4.

For 5-FU imaging experiments, 5-FU was delivered via retro-orbital injection as a single dose of 150 mg/kg as described above. In vivo bone marrow two-photon imaging was then performed on day 4 or day 20 after the 5-FU injection. The calvarial cell location map in Extended Data Fig. 8c was made in similar way to the Cy/GCSF cell maps described above.

For Cy/GCSF ex vivo imaging, freshly fixed (4% paraformaldehyde) and excised mouse calvaria were affixed to a plastic dish, immersed in $1\times$ PBS, and immediately imaged for 4–5 h. Tiled z-stacks were acquired with $3\text{-}\mu\text{m}$ steps and a 10% overlap between fields of view. Images were stitched together in 3D using Olympus FluoView software or ImageJ scripts.

For examining bone remodelling activities in calvaria (in vivo) and metaphysis (ex vivo), calcium binding dyes were administered 48 h apart via retro-orbital injection. Calvarial in vivo imaging was performed as described. Mouse tibia was freshly removed, thinned, and imaged from the bone surface. Tiled z-stacks were acquired with $3\text{-}\mu\text{m}$ steps and stitched using ImageJ.

Image quantification

For distance measurements, the distance from each cell to blood vessels or to the nearest bone surface (that is, endosteum identified using SHG) was computed by hand as described previously using the Pythagorean theorem^{27,42}. The bone contains abundant collagen, which enabled us to use SHG imaging to identify the inner bone surface (endosteum). This technique has been used in many previous publications of live bone marrow imaging^{3,27,43,44}.

The identity of blood vessels (arterioles, sinusoids, or transition vessels) within the calvaria was determined by a combination of morphology, location within the vessel network, location within the bone marrow, and blood flow. In brief, with our blood pool agent the arterioles appear as narrow ($\sim 5\text{--}10\ \mu\text{m}$ diameter) and generally straight vessels with a smooth surface upstream of sinusoidal vessels, which are larger ($\sim 20\ \mu\text{m}$ diameter or greater) with irregular surfaces. This definition was based on previous work^{27,44}, which confirmed that these small diameter vessels are arterioles with a faster flow speed ($\sim 2\text{ mm/s}$ or higher), higher pO_2 , and increased barrier function in comparison to sinusoidal vessels. They also stain positive for SCA1. At the transition point between arterioles and sinusoids, the vessel diameter increases. It is from this point of increase to the next vessel branching point downstream that we define as transitional vessels.

Distance measurements were performed in ImageJ v.1.51p. For display purposes, the brightness and contrast of images in the figures were adjusted, but all image analysis was performed on raw data. For motility measurements, frame-to-frame drift was corrected in 3D using the Template Matching plugin in ImageJ. Next, the centroid of the cell was determined for the first and last image of a 20-min sequence, and the 2D displacement was calculated using the distance formula.

For cell clustering analysis, the individual tiled z-stack images were reconstructed into a single z-stack for the whole calvaria using ImageJ. Next, each cell was designated as one of three tags (red, green, or blue) based on the colour of the cell during imaging, and the x, y, z coordinates were recorded by hand. Using a custom Matlab script similar to ClusterQuant⁴⁵, we analysed the spatial clustering (cluster size = 3) of like-coloured cells in this model compared to 10,000 randomized samples to determine the statistical likelihood of the colour clustering in our samples. Graphs and statistical analyses were performed using Graph Pad Prism version 6 or higher. The contrast

and/or brightness of figure images and videos were adjusted for display purposes only.

Classification of bone marrow cavities

A bone marrow cavity is defined as a 3D inclusion inside bone with a single concave endosteum (Fig. 4e, Extended Data Fig. 11d–f), while deeper down all cavities are interconnected. Once a cavity has been defined using the bone SHG signal, we classified types of bone marrow cavity by sequential staining with two calcium-binding reagents. The first calcium-binding dye (dye 1, tetracycline or calcein blue, Sigma; 35 mg/kg and 30 mg/kg, respectively) was administered 48 h before imaging to track bone resorption activities based on erosion of dye 1, and the second calcium-binding dye (dye 2, Alizarin red, 40 mg/kg) was administered 30 min before imaging to label high-calcium regions (bone fronts). The 48-h interval was chosen on the basis of the estimated lifespan of mouse active osteoclasts^{46,47}. Therefore, the double-staining approach delineated approximately one bone erosion cycle in the bone marrow. As the lack of dye 1 indicates the existence of resorption whereas strong double staining with both dyes indicates ongoing bone deposition, the dye 1:dye 2 ratio contained within a single concave endosteum depicts the status of bone remodelling during the 48-h period. For each cavity, the acquired depth covered the dye 1- and dye 2-labelled regions, typically between 80 and 120 μm beneath the endosteum. For quantification of osteoblast or osteoclast coverage (2.3Col1GFP or cathepsin K pixels) along the endosteum, z-stack double staining, COL1, and cathepsin K images of $2\text{-}\mu\text{m}$ z-steps were rendered in 2D using maximum intensity projection in ImageJ and then analysed (Extended Data Fig. 9). As cathepsin K is also expressed substantially by endothelial cells, a vascular map (rhodamine B dextran) was acquired simultaneously and subtracted from the cathepsin K map before we retrieved the total pixel counts. For quantifying fractions of cavity types, 3D maps of calvaria were acquired and rendered in 2D using maximum intensity projection, then analysed (Fig. 4f). For quantification of cavity types for Fig. 4g, h, total pixels of dye 1 and dye 2 were retrieved directly from the 3D stacks. Segmentation of dye 1 and dye 2 in each stack was obtained using ImageJ macros combining multiple built-in plugins. Specifically, contrast enhancement was applied consistently (0.1% saturation) for each stack. The images were smoothed using 3D imageJ suite plugins⁴⁸ (3D mean filter, kernel size = 1) followed by background subtraction using the rolling ball algorithm with radius size of 100 and 250 pixels for dye 1 and dye 2, respectively. This background subtraction step removed diffuse signals from bone autofluorescence that are more prominent in blue and green channels (dye 1) but still distinct from structured patterns of bone front staining. Segmentation was then performed using ImageJ built-in global or local thresholding algorithms to render matching binary results compared to raw stacks. The total numbers of dye 1 and dye 2 pixels were then obtained from the binary images to calculate the dye 1:dye 2 ratio.

For classification purposes, we defined bone cavities as (i) deposition type (D-type; dye 1:dye 2 > 75%); (ii) resorption type (R-type; dye 1:dye 2 < 25%), and (iii) mixed type (M-type; dye 1:dye 2 25–75%). These definitions emphasize functional perspectives of bone remodelling along the endosteum (dominated by bone deposition or resorption), instead of the presence of osteoblasts or osteoclasts at the time of imaging. This is especially important given that osteoclasts went through apoptosis after each resorption cycle and may not be present at the time of imaging. Of note, bone-lining cells have been reported to occupy the bone fronts of inactive regions that lack both mature osteoblasts and calcium staining⁴⁹. In the calvarium, as neither MDS-HSPCs nor MFG-HSCs were found in fully inactive regions, we characterized cell distribution only in D, M, and R cavities, where small patches of inactive areas may be present but do not alter the distribution of the three cavity types. To quantify the number of cells per bone marrow cavity (Fig. 4g, h), *Mds1*^{GFP/+}*Flt3*^{Cre} and *Mds1*^{GFP/+} cells were manually counted. A cell was

considered to belong to a cavity if it was underneath a concave dome; there was no restriction on its distance from the endosteal surface.

Bone clearing and imaging of femurs

Tissue preparation, multicolour full-bone imaging of thick femoral sections and quantitative analysis were performed as previously described²⁴. In brief, bones were fixed for 18 h in 4% paraformaldehyde before being decalcified using 10% EDTA (EDTA, pH = 8) for two weeks. Longitudinal bone sections (250 μ m thick) were blocked, permeabilized (followed by additional blocking of endogenous avidins and biotins) and stained overnight at room temperature with primary antibodies (anti-GFP (chicken, Aves Labs, GFP-1020), anti-CD117 (goat, R&D, AF1356), anti-CD105 (rat, eBioscience, 14-1051-82) and anti-collagen type I (rabbit, Cedarlane, CL50151AP)), secondary antibodies (Alexa Fluor 555, 680, and CF633) and DAPI (Thermo Fischer Scientific, D1306). The GFP signal was amplified using donkey anti-chicken biotin (Jackson ImmunoResearch, 703-065-155) followed by streptavidin Alexa 488 (Thermo Fischer Scientific, S32354). Full-bone scans were performed using a Leica TCS SP8 confocal microscope equipped with three photomultiplier tubes and two HyD detectors using type F immersion liquid (RI: 1.518) and a 20 \times multiple immersion lens (NA 0.75, FWD 0.680 mm). Images were acquired at 8-bit, 400 Hz and 1,024 \times 1,024 resolution with 2.49 μ m z-spacing. Segmentation and distance quantification analysis were performed with Imaris (version 8.3.1), using the XT and Distance Transformation XTension modules. To avoid data truncation, data were transformed to 16 bit before distance quantification and then reverted back to 8 bit. Random dots were generated via a Matlab-based, self-developed software (XiT) as previously described²⁴. Graphs and statistical analyses were performed using Graph Pad Prism version 6.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The GEO accession number for GFP cells is GSE115908. The GEO accession number for LSK cells used for overlay has been published previously (GSE90742)¹⁸.

35. Rodríguez, C. I. et al. High-efficiency deleter mice show that FLP is an alternative to Cre-loxP. *Nat. Genet.* **25**, 139–140 (2000).
36. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
37. Kalajic, Z. et al. Directing the expression of a green fluorescent protein transgene in differentiated osteoblasts: comparison between rat type I collagen and rat osteocalcin promoters. *Bone* **31**, 654–660 (2002).
38. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).

39. Esipova, T. V. et al. Two new “protected” oxyphors for biological oximetry: properties and application in tumor imaging. *Anal. Chem.* **83**, 8756–8765 (2011).
40. Lebedev, A. Y. et al. Dendritic phosphorescent probes for oxygen imaging in biological systems. *ACS Appl. Mater. Interfaces* **1**, 1292–1304 (2009).
41. Esipova, T. V., Rivera-Jacquez, H. J., Weber, B., Masunov, A. E. & Vinogradov, S. A. Two-photon absorbing phosphorescent metalloporphyrins: effects of π -extension and peripheral substitution. *J. Am. Chem. Soc.* **138**, 15648–15662 (2016).
42. Lo Celso, C., Wu, J. W. & Lin, C. P. In vivo imaging of hematopoietic stem cells and their microenvironment. *J. Biophotonics* **2**, 619–631 (2009).
43. Bixel, M. G. et al. Flow dynamics and HSPC homing in bone marrow microvessels. *Cell Rep.* **18**, 1804–1816 (2017).
44. Itkin, T. et al. Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* **532**, 323–328 (2016).
45. Mondor, I. et al. Clonal proliferation and stochastic pruning orchestrate lymph node vasculature remodeling. *Immunity* **45**, 877–888 (2016).
46. Manolagas, S. C. Birth and death of bone cells: basic regulatory mechanisms and implications for the pathogenesis and treatment of osteoporosis. *Endocr. Rev.* **21**, 115–137 (2000).
47. Weinstein, R. S. et al. Promotion of osteoclast survival and antagonism of bisphosphonate-induced osteoclast apoptosis by glucocorticoids. *J. Clin. Invest.* **109**, 1041–1048 (2002).
48. Ollion, J., Cochenec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).
49. Matic, I. et al. Quiescent bone lining cells are a major source of osteoblasts during adulthood. *Stem Cells* **34**, 2930–2942 (2016).

Acknowledgements We thank Camargo laboratory members for discussions, D. Scadden and N. Severe for discussions and help with long bone staining, L. Kunz for generating initial Matlab scripts facilitating preliminary long bone data analysis and R. Matthieu of the Stem Cell Program Flow Cytometry facility for FACS assistance. This study was supported by awards from the National Institute of Health (HL128850-01A1 and P01HL13147 to F.D.C., R01EB017274, R01CA194596, R01DK115577, and R24DK103074 to C.P.L., R01EB018464, R24NS092986, EB018464 and NS092986 to S.A.V., R01CA175761 to A.S.P., and NIDDK-supported Cooperative Centers of Excellence in Hematology at BCH U54DK110805). F.D.C. is a Leukemia and Lymphoma Society Scholar and a Howard Hughes Medical Institute Scholar.

Author contributions C.C. and F.D.C. designed experiments relevant to the animal models. J.A.S. and C.P.L. designed experiments relevant to live animal calvaria bone marrow imaging and fixed calvaria imaging. S.-C.A.Y. and C.P.L. designed experiments relevant to imaging of bone cavity types in the calvaria and tibia. K.D.K. and T.S. designed experiments relevant to femur staining and imaging. C.C., A.R., A.S.P., Y.Z. and S.R. generated the mouse models. C.C. performed all animal-related experiments and relevant data analysis. R.A.C. and R.P. supervised and performed the bioinformatics analysis, respectively. J.A.S. and N.S. performed the live animal calvaria imaging experiments, fixed calvaria imaging, and relevant data analysis. R.T. performed part of the live animal calvaria imaging experiments and relevant data analysis. T.V.E. and S.A.V. generated the pO₂ probe and performed relevant characterization. K.D.K. performed the long bone imaging experiments and data analysis. S.-C.A.Y. performed the imaging experiments and analysis of bone cavity types and cell proliferation. S.H.O. and G.G. designed the fluidigm experiments. G.G. performed the fluidigm experiments and related analysis. C.C., J.A.S., S.-C.A.Y., C.P.L. and F.D.C. wrote the manuscript. C.P.L. and F.D.C. supervised the project and gave final approval.

Competing interests The authors declare no competing interests.

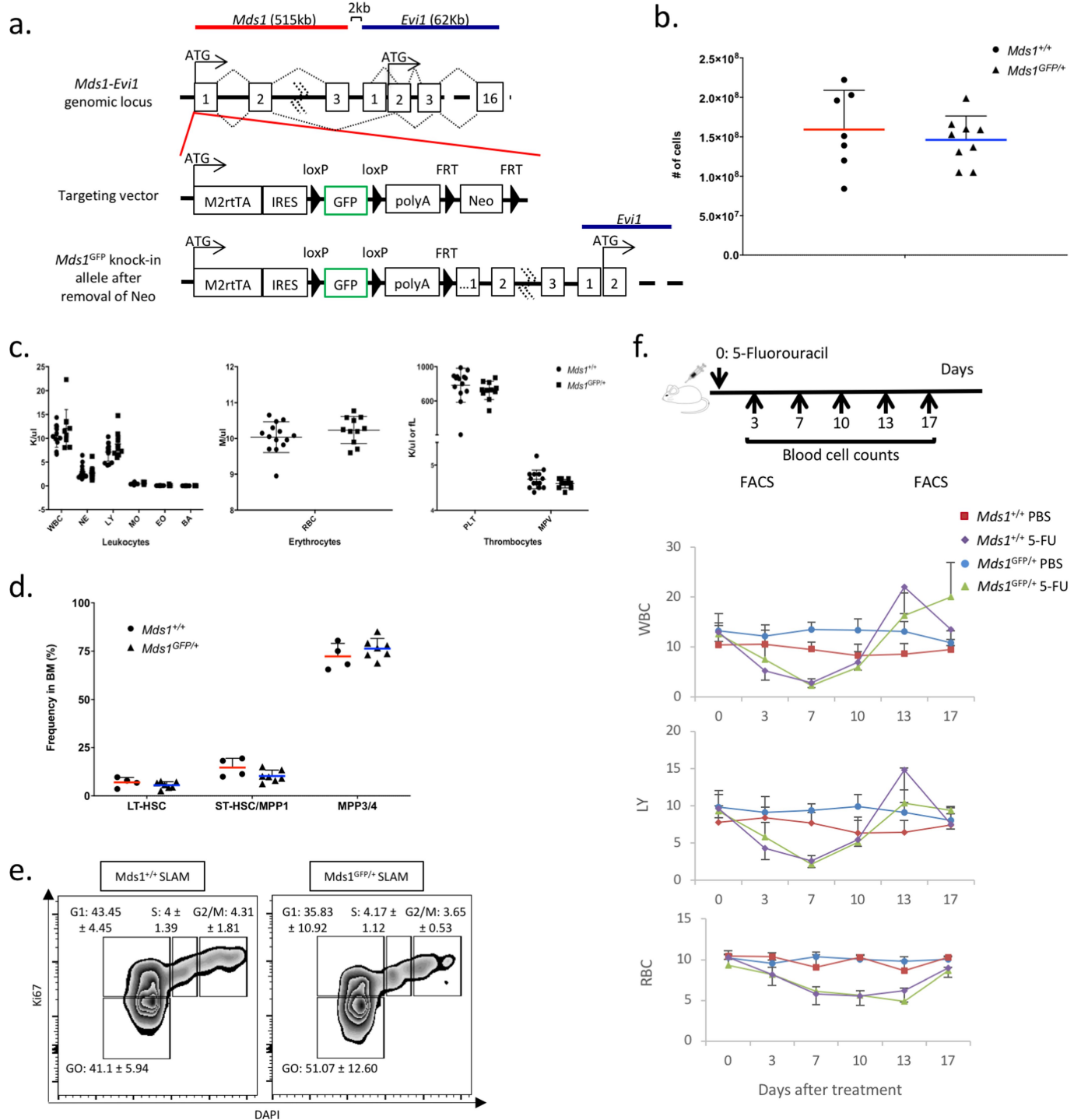
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1971-z>.

Correspondence and requests for materials should be addressed to C.P.L. or F.D.C.

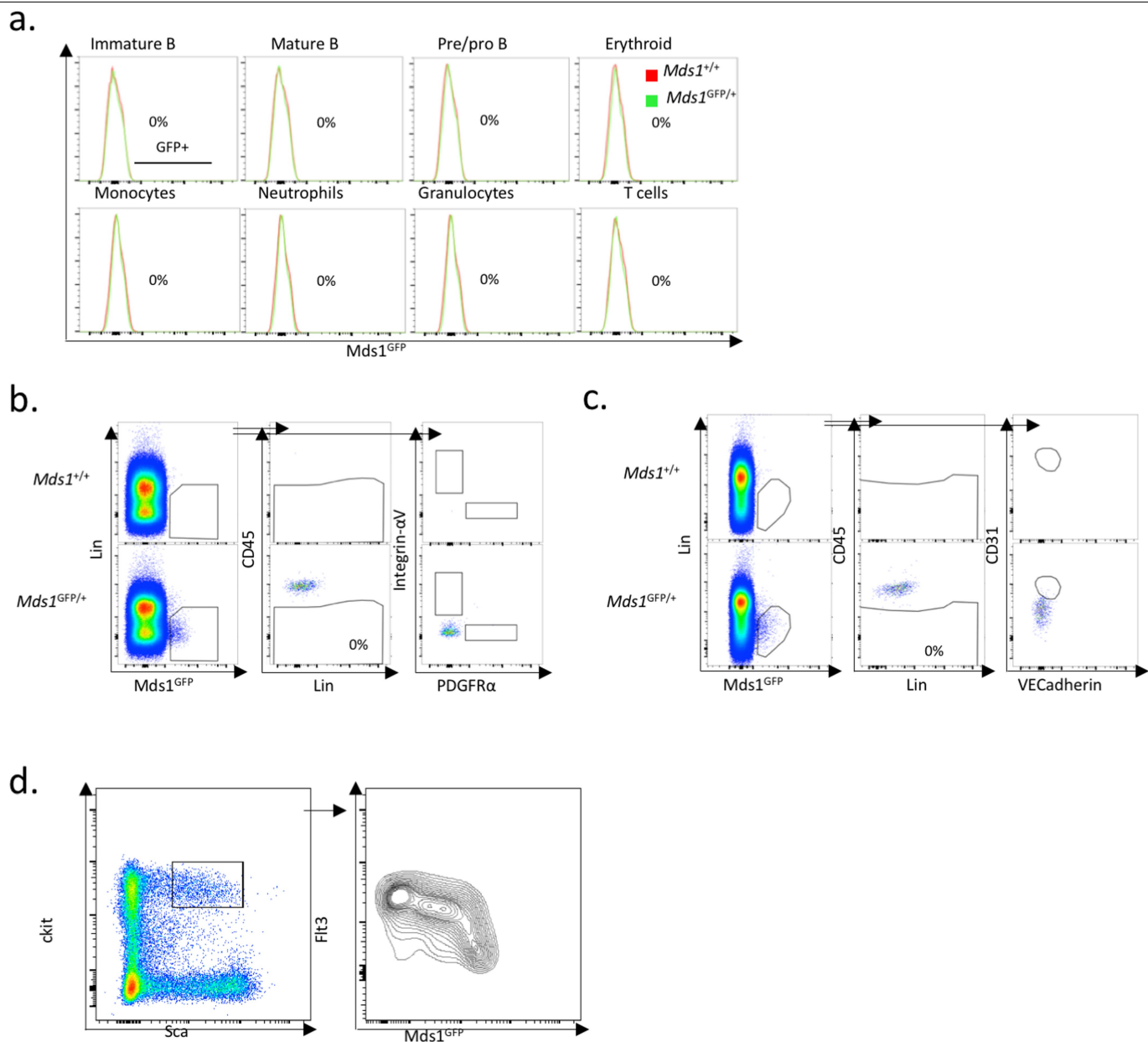
Peer review information Nature thanks Elisa Laurenti, Cristina Lo Celso and Hanna Mikkola for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Characterization of HSPC (*Mds1^{GFP/+}*) mice demonstrates normal haematopoiesis, HSC frequency, cell cycle and stimuli recovery response. **a, Targeting strategy for the generation of *Mds1^{GFP/+}* mice. **b**, Eight-to-twelve-week-old *Mds1^{GFP/+}* mice ($n=9$) show similar bone marrow cellularity to control mice (*Mds1^{+/+}*; $n=7$); mean \pm s.d. **c**, eight-to-twelve-week-old *Mds1^{GFP/+}* mice ($n=14$) have similar peripheral blood parameters to *Mds1^{+/+}* control mice ($n=11$); mean \pm s.d. **d**, Eight-to-twelve-week-**

old *Mds1^{GFP/+}* mice ($n=7$) showed similar frequencies of CD150⁺CD48⁺LSK (LT-HSCs), CD150⁺CD48⁺LSK (ST-HSCs) and CD150⁺CD48⁺LSK (MPPs) to control mice (*Mds1^{+/+}*; $n=4$); mean \pm s.d. **e**, Cell cycle analysis of SLAM cells from *Mds1^{GFP/+}* ($n=3$) and wild-type (*Mds1^{+/+}*; $n=2$) mice in native conditions. Indicated value per gate represents mean \pm s.d. **f**, Dynamics of recovery of white blood cells (WBC), lymphocytes (LY) and red blood cells (RBC) upon 5-FU treatment in *Mds1^{GFP/+}* and control (*Mds1^{+/+}*) mice. Mean \pm s.d., $n=4$ mice.

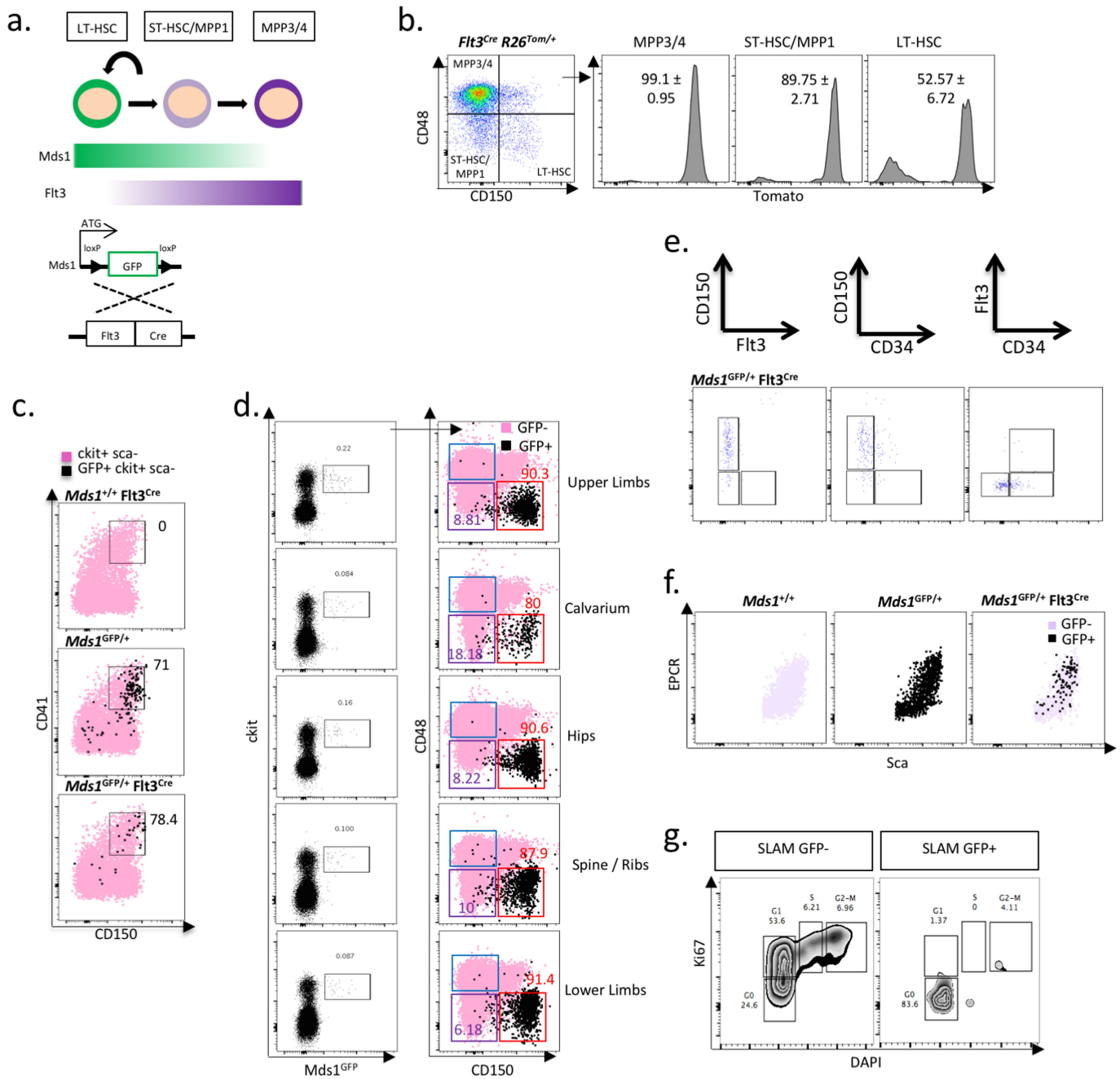


Extended Data Fig. 2 | Flow cytometric analysis of *Mds1*^{GFP/+} expression.

a, GFP⁺ cells are not present in any mature cellular subpopulations. Data shown are from one representative experiment that was repeated three times.

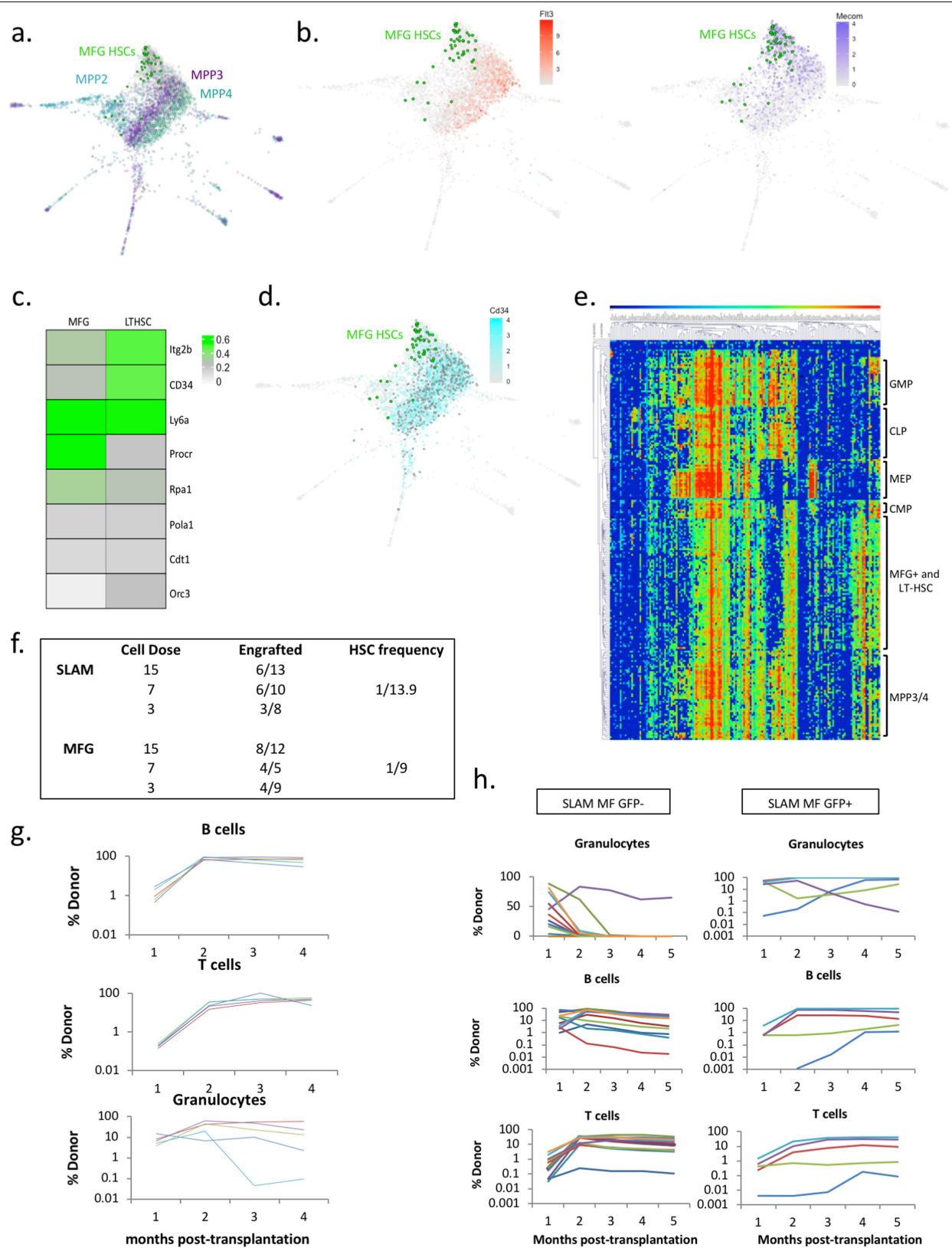
b, c, *Mds1*^{GFP/+} cells are not present in the CD45⁺ bone marrow compartment or in mesenchymal (integrin- α V and PDGFR α) or endothelial (CD31 and

VE-cadherin) bone marrow niche components. The experiment was performed once. **d**, Flow cytometry analysis reveals an inverse correlation between MDS1-GFP expression and FLT3 staining in LIN⁺ SCA⁺ CKIT⁺ cells. The experiment was performed twice with similar results.



Extended Data Fig. 3 | Generation of the MFG (*Mds1^{GFP/+} Flt3^{Cre}*) mice results in restriction of GFP expression to LT-HSCs. **a, Schematic of genetic strategy to restrict GFP expression to LT-HSC compartment. **b**, Analysis of bone marrow from *Flt3^{Cre} R26^{LSL-Tom}* mice shows *Flt3^{Cre}*-driven activity in compartments downstream of LT-HSCs. *n* = 4 mice; mean ± s.d. **c**, Further characterization of the CKIT⁺SCA1⁺GFP⁺ cells from MFG mice. CD41⁺CD150⁺ cells represent megakaryocyte progenitors. The experiment was performed twice with similar results. **d**, Flow characterization of MFG cell in marrow**

isolated from multiple bones. The experiment was performed three times with similar results. **e**, MFG-HSCs are predominantly found within the CD34⁺Flt3⁺CD150⁺ bone marrow fraction. The experiment was performed twice with similar results. **f**, MFG-HSCs are predominantly found within the SCA1^{high}EPCR⁺ bone marrow fraction. The experiment was performed once. **g**, Cell cycle analysis of SLAM cells that are either GFP⁺ or GFP⁻ in MFG mice. Pooled data from three mice.

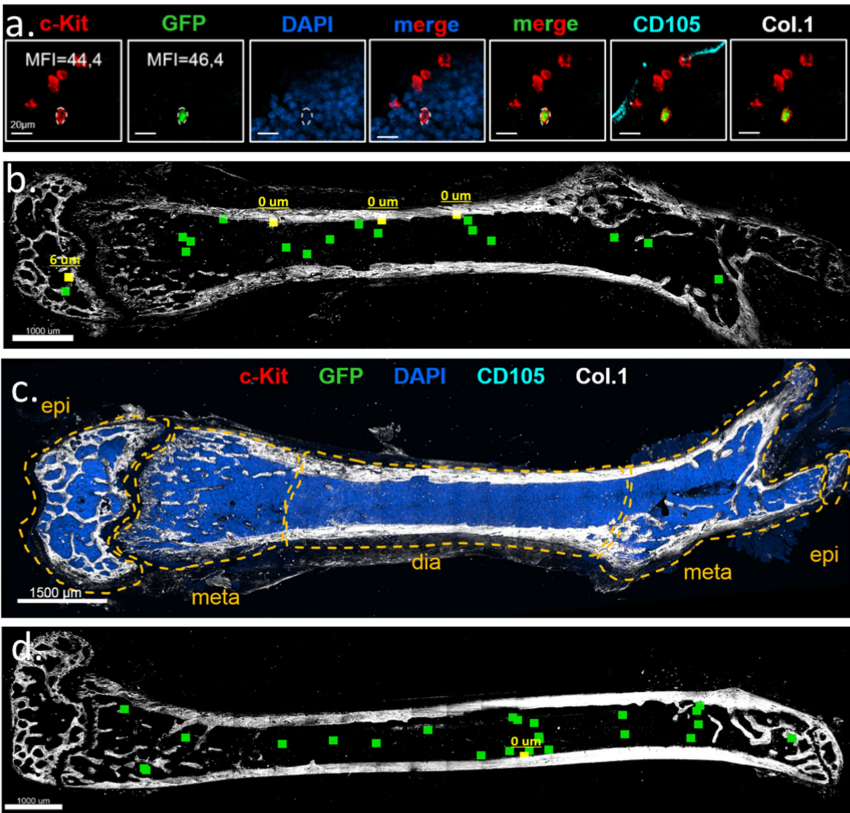


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Additional characterization of MFG-HSCs.

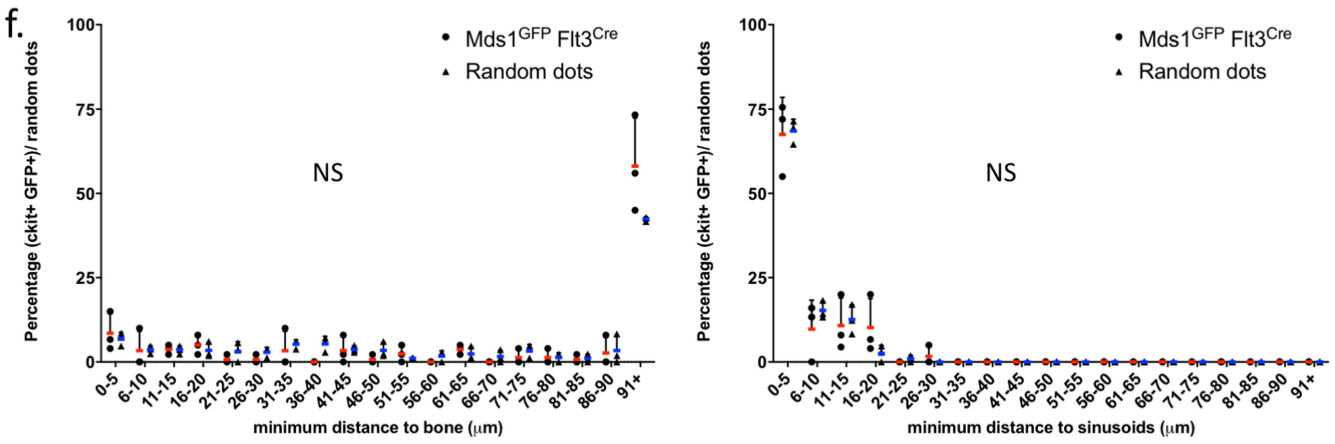
a, b, InDrops scRNA-seq analysis of MFG⁺ cells in comparison to multiple populations of HSC and MPPs. MFG cells (46 cells) are predominantly found in areas where *Mecom* (purple, $n = 742$ cells), but not *Flt3* (orange, $n = 1,111$ cells) is expressed. Teal, MPP2; purple, MPP3; light green, MPP4; grey, other cells; bright green, *Mds1*^{GFP/+} *Flt3*^{Cre} cells. Gradient colour demonstrates normalized read counts. Each dot represents an individual cell. MFG-HSCs represent cells from a single mouse, the rest of the cells represent cells from a separate single mouse. **c, d**, Heatmaps (**c**) and spring plot map (**d**) showing expression levels of previously published 'dormant' HSC genes in scRNA-seq data from LTHSC and MFG cell populations. For the spring plot analysis: MFG, $n = 46$ cells; CD34, $n = 2,380$ cells (teal); each dot represents an individual cell. MFG-HSCs represent cells from a single mouse; the rest of the cells represent cells from a

separate single mouse. **e**, Single-cell transcriptional fluidigm profile of MFG-HSCs demonstrates that they cluster together with LT-HSCs. **f**, Summary of transplants with 3, 7, or 15 MFG or SLAM HSCs together with 100,000 bone marrow cells, analysed 4 months after transplantation. HSC frequencies were calculated using ELDA software (see Methods). **g**, Engraftment analysis following secondary transplantations using whole bone marrow from one primary recipient of 25 MFG⁺ HSCs. Experiment shown is representative of three independent experiments. **h**, Percentage chimaerism at 4, 8, 12, 16 and 20 weeks in primary recipients transplanted with 25 SLAM cells sorted on the basis of GFP expression isolated from *Mds1*^{GFP/+} *Flt3*^{Cre} mice ($n = 12$ GFP⁺ mice, $n = 5$ GFP⁻ mice). Our data demonstrate that GFP⁺ cells within the SLAM compartment are more functionally enriched. Each line represents an individual mouse.



e.

	No of c-kit+ GFP+ cells	Epiphysis	Metaphysis	Diaphysis
N=1	25	0 (0%)	9 (36%)	16 (64%)
N=2	45	3 (7%)	14 (31%)	28 (62%)
N=3	20	2 (10%)	6 (30%)	12 (60%)
Average	30	1.67	9.67	18.67



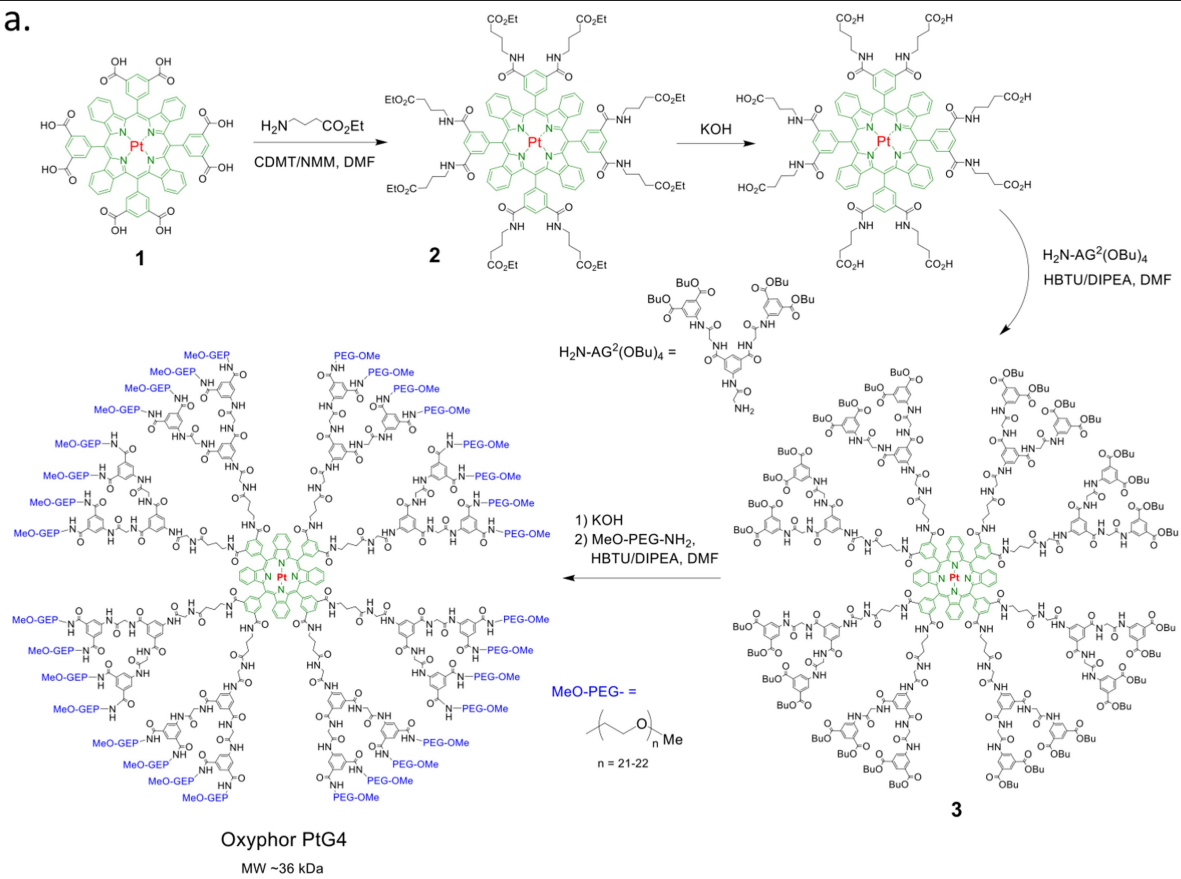
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Multicolour quantitative deep-tissue confocal imaging of complete femoral sections from MFG (*Mds1^{GFP/+}Flt3^{Cre}*) mice.

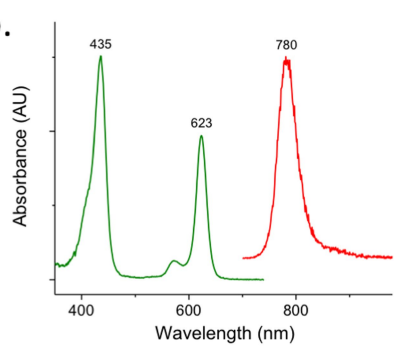
a, Identification of C-KIT⁺GFP⁺MFG-HSCs using multicolour quantitative deep-tissue confocal imaging of full bone femoral sections. Pictures are 10- μ m.xy projections of one area of interest. $n = 3$ mice. The experiment was performed three times with similar results. **b**, Example of one full-bone femoral section with colour-coded visualization of HSCs based on their distance from bone. Yellow squares represent individual HSCs in proximity to cortical or trabecular bone, whereas green dots represent individual HSCs located more than 10 μ m away. The picture represents data from an individual mouse. The experiment was performed three times with similar results (**d**). **c**, Example of full-bone femoral section (only Col.1 and DAPI staining are shown). The experiment was performed three times with similar results. **d**, Colour-coded visualization of HSCs based on their distance to bone. Yellow squares represent individual

HSCs in proximity to cortical or trabecular bone, whereas green dots represent individual HSCs located more than 10 μ m away. This picture represents an independent mouse from **b**. The experiment was performed three times with similar results. **e**, Quantification of absolute number and anatomical location of C-KIT⁺GFP⁺MFG-HSCs per individual experiment. ($N = 3$ mice) **f**, Spatial distribution of HSCs (circles) and random dots (triangles) relative to Col.1 marking bone surfaces (left panel) and CD105⁺ vasculature (sinusoids, right panel) ($n = 3$ mice). P values were calculated using two-tailed Kolmogorov–Smirnov (distance distributions, left panel $P = 0.1516$, right panel $P > 0.9999$) and one-tailed Mann–Whitney (first bin of histograms, left, HSCs: 8.56 ± 5.74 , RDs: 6.88 ± 1.94 , $P = 0.50$; right, HSCs: 67.52 ± 10.99 , random dots: 68.53 ± 3.51 , $P = 0.35$) tests. Data points with mean \pm s.d. (red for HSCs, blue for random dots). NS, not significant. Epi: epiphysis, meta: metaphysis, dia: diaphysis.

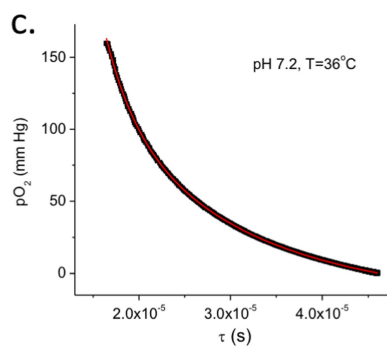
a.



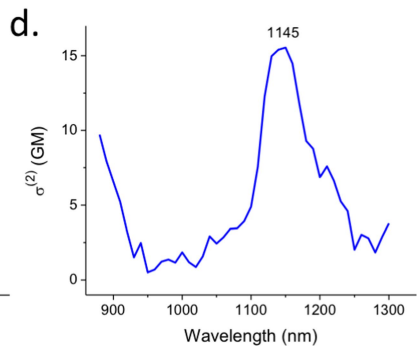
b.



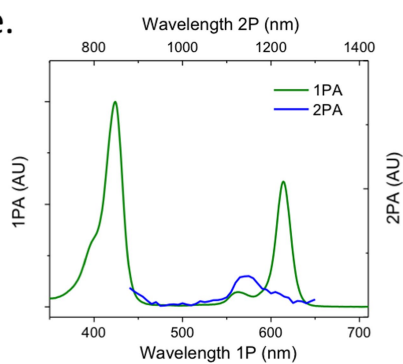
c.



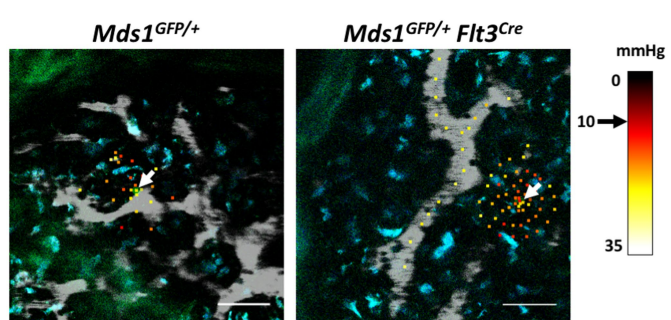
d.



e.



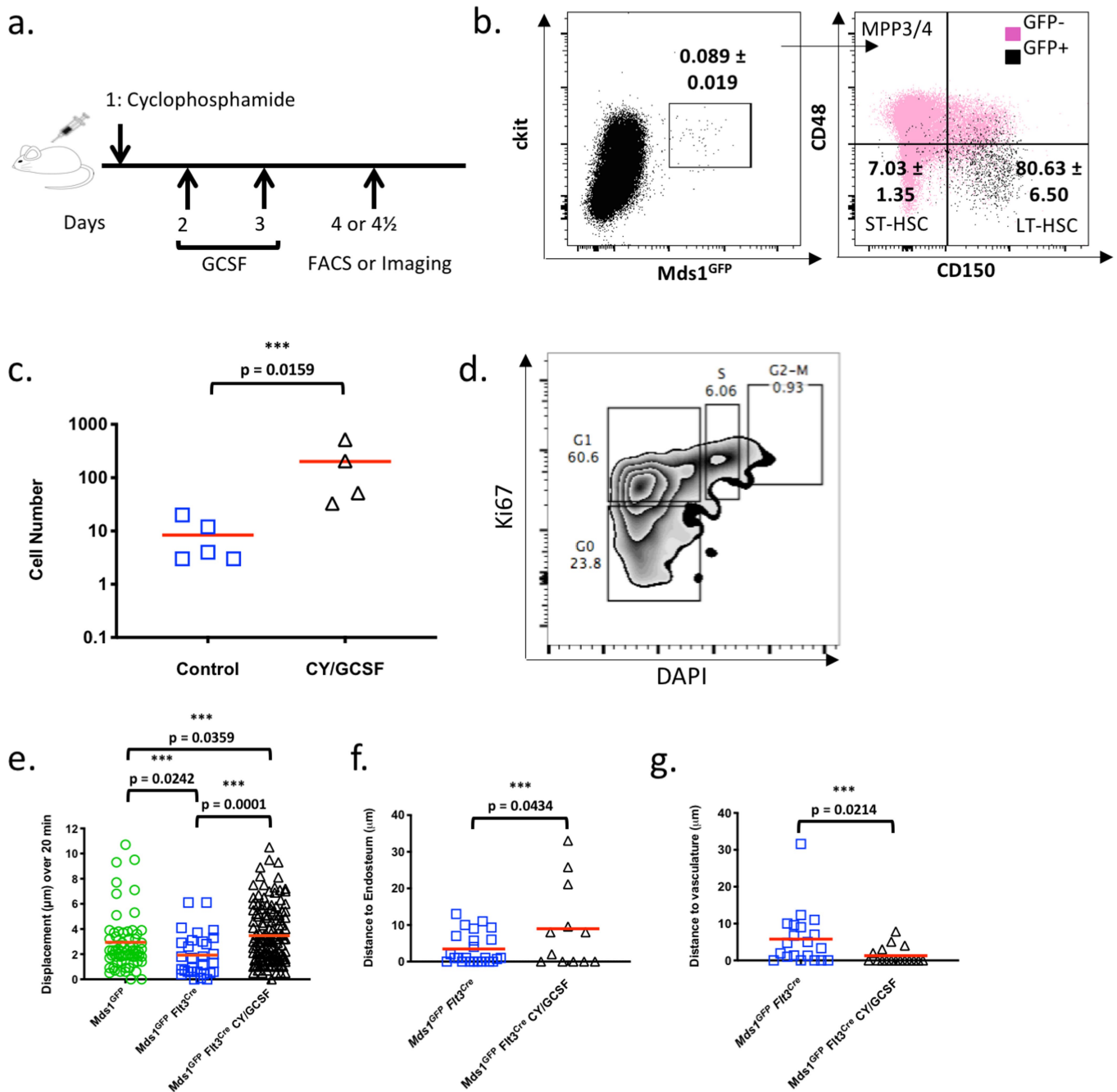
f.



Extended Data Fig. 6 | See next page for caption.

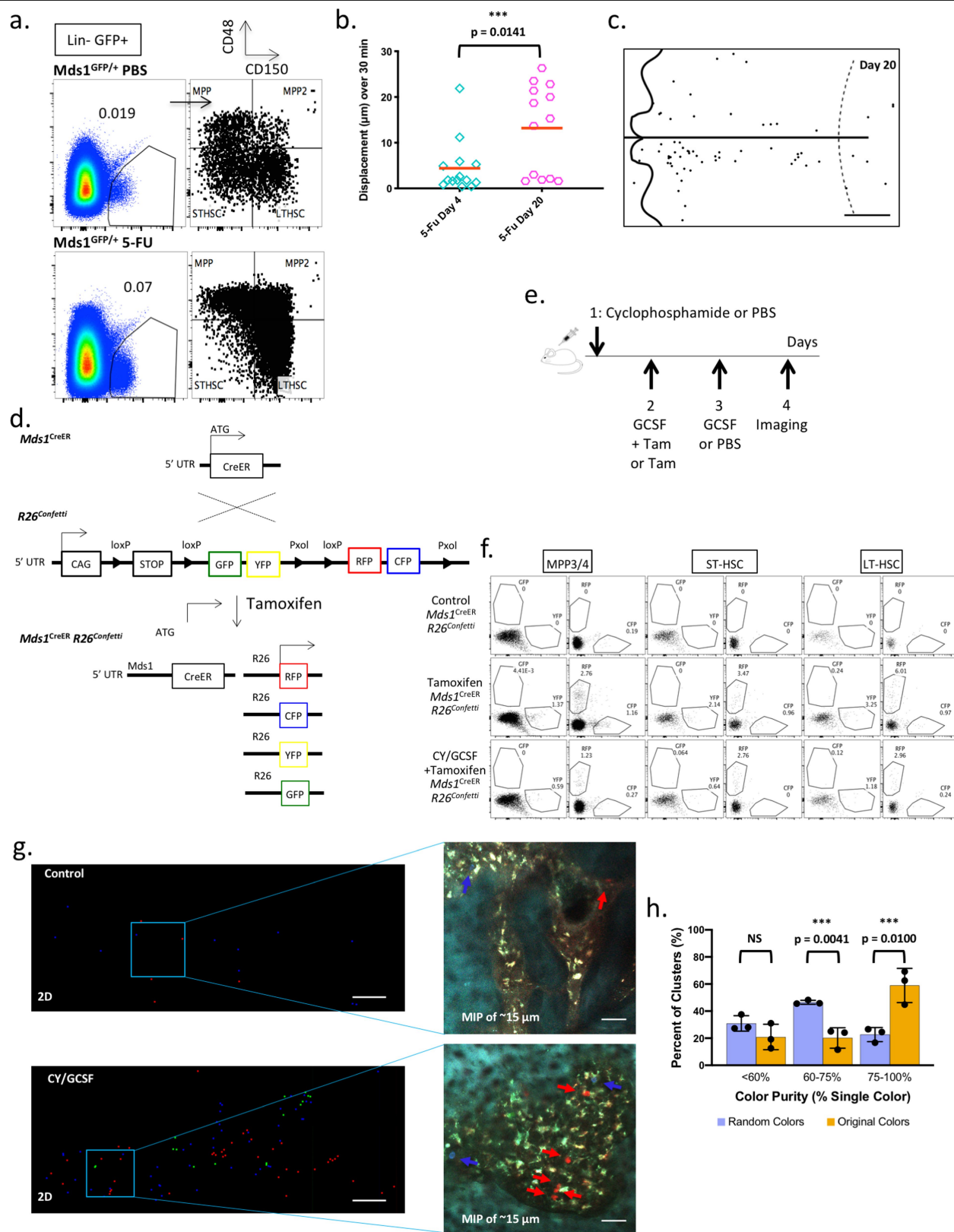
Extended Data Fig. 6 | Synthesis, structure and characterization of phosphorescent probe Oxyphor PtG4. The structure of Oxyphor PtG4 is almost identical to that of the previously published probe Oxyphor PdG4³⁹, but it contains Pt instead of Pd at the core of the porphyrin (**1**: Pt *tetra-meso*-3,5-dicarboxyphenyl-tetrabenzoporphyrin). **a**, Synthesis of Oxyphor PtG4. First, eight carboxyl groups on the porphyrin **1** were amended with 4-amino-ethylbutyrate linkers. Upon hydrolysis of the terminal esters in the resulting porphyrin **2**, eight aryl-glycine dendrons ($\text{H}_2\text{N-AG}^2(\text{OBu})_4$) were coupled to the resulting porphyrin-octacarboxylic acid, giving dendrimer **3**. The butyl esters on the latter were hydrolysed under mild basic conditions, and the resulting free carboxylic acid groups were amidated with mono-methoxypolyethyleneglycol amine (MeO-PEG-NH_2 , Av MW 1000), giving the target probe Oxyphor PtG4. MALDI-TOF (m/z) was used to confirm the identity of the intermediate products as well as of the target probe molecule. Structure **2** ($\text{C}_{116}\text{H}_{124}\text{N}_{12}\text{O}_{24}\text{Pt}$, calculated at MW 2,263.85) was found 2,264.48 $[\text{M}]^+$; structure **3** ($\text{C}_{468}\text{H}_{540}\text{N}_{60}\text{O}_{120}\text{Pt}$, calculated at MW 9,114.76) was found at 9,115.68 $[\text{M}+\text{H}]^+$ and Oxyphor PtG4 ($\text{C}_{1780}\text{H}_{3196}\text{N}_{92}\text{O}_{792}\text{Pt}$, calculated at MW 40,538) was found at 35,952. For Oxyphor PtG4 we identified an additional peak at MW 66,123.6 which is probably due to the presence of dimeric species formed during the ionization process. **b**, Linear (one photon) absorption (green) and emission spectra (red) of PtG4 in 50 mM phosphate buffer solution (pH 7.2, $\lambda_{\text{ex}} = 623 \text{ nm}$; photophysical constants in PBS, 22 °C: $\epsilon(623) \sim 90,000 \text{ M}^{-1} \text{ cm}^{-1}$

(molar extinction coefficient), $\phi_{\text{phos}}(\text{deox}) \sim 0.07$ (phosphorescence quantum yield in deoxygenated solution), $\tau_{\text{air}} = 16 \mu\text{s}$ (phosphorescence decay time on air), $t_{\text{deox}} = 47 \text{ ms}$ (phosphorescence decay time in deoxygenated solution). **c**, Phosphorescence oxygen quenching plot of Oxyphor PtG4. The calibration was performed as previously described³⁹. The experimental points were fitted to an arbitrary double-exponential form and the obtained parametric equation was used to convert the phosphorescence lifetimes obtained in vivo experiments to pO_2 values. **d**, Two-photon absorption spectrum of PtG4 in deoxygenated dimethylacetamide (DMA, 22 °C). **e**, Arbitrarily scaled one- (green line) and two-photon (blue line) absorption spectra of PtG4. The two-photon absorption (2PA) spectra of PtG4 and of the reference compounds were measured by the relative phosphorescence method, as previously described⁴¹. The laser source was a Ti:Sapphire oscillator (80 MHz rep. rate) with tunability range of 680–1,300 nm (Insight Deep See, Spectra Physics). All optical spectroscopic experiments and oxygen titrations were performed at least three times, giving highly reproducible results. **f**, Representative intravital images of an HSPC (green, left image), MFG-HSC (green, right image), vasculature (grey, Rhodamine-B-dextran 70 kDa), and autofluorescence (blue) overlaid with localized oxygenation measurements. White arrows, GFP cells. Black arrow, colour representing 10 mm Hg. Coloured squares represent individual localized oxygen measurement areas. Images represent data from two independent experiments for each mouse model. Scale bars, $\sim 50 \mu\text{m}$.



Extended Data Fig. 7 | Increased motility and expansion of activated MFG-HSCs. **a.** Schematic illustration of protocol for activating bone marrow HSCs using Cy/G-CSF. **b.** Flow cytometry analysis of Cy/G-CSF-treated MFG mice ($n = 3$ mice). Data show Lineage⁻ cells. Mean \pm s.d. **c.** Number of GFP⁺ cells identified per calvaria in untreated and Cy/G-CSF-treated *Mds1^{GFP/+} Flt3^{Cre}* mice ($n = 5$ and 4 mice, respectively). Red bars indicate mean. P was calculated using two-tailed Mann-Whitney test. **d.** Cell cycle analysis of MFG⁺ cells from Cy/G-CSF-treated mice. Three mice were pooled together to acquire the displayed data. **e.** Graph

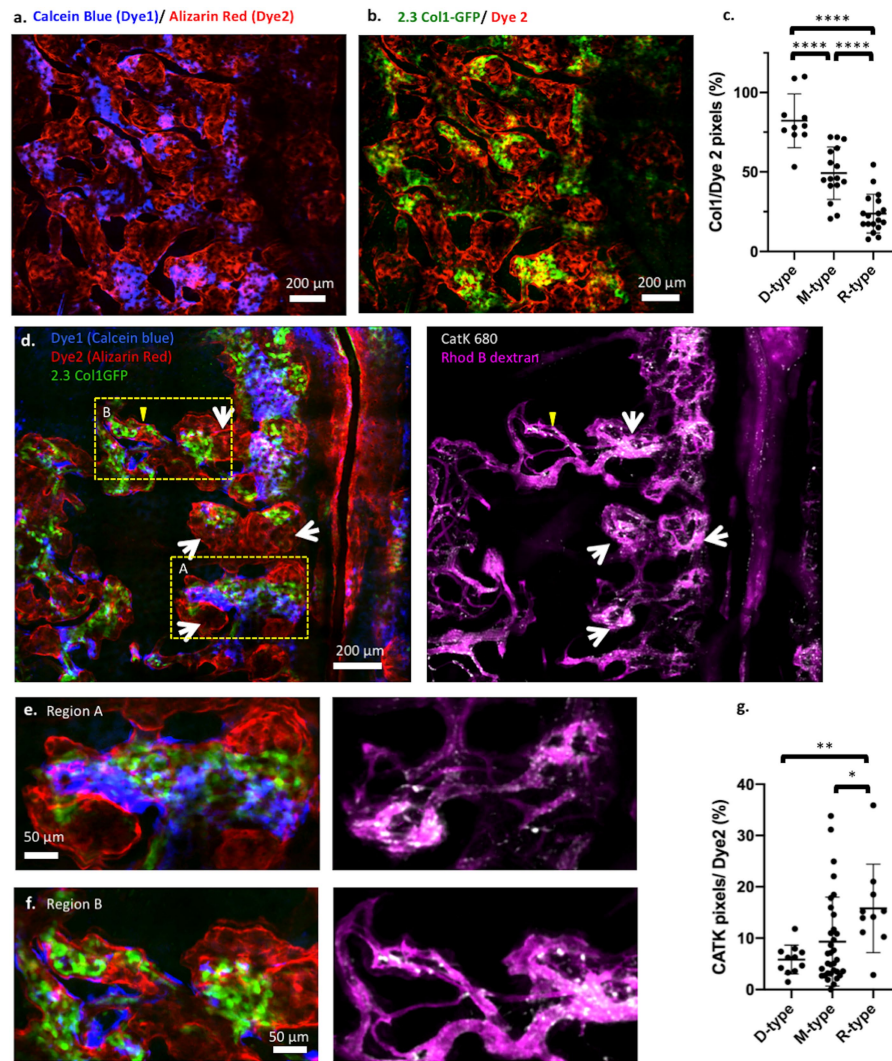
showing in vivo motility measurements of HSPCs ($n = 66$ cells) and MFG-HSCs ($n = 30$ cells) at steady-state and of activated MFG-HSCs ($n = 142$ cells) in the calvaria. Red bars indicate mean. P were calculated using two-tailed Mann-Whitney test. **f, g.** Distance from MFG⁺ cells to the endosteum ($n = 24$ and 12 cells for untreated and Cy/G-CSF-treated, respectively) and to the nearest vessel ($n = 20$ and 17 cells for untreated and Cy/G-CSF-treated, respectively), after treatment with Cy/G-CSF. Red bars indicate mean. P values calculated using two-tailed unpaired t -test.



Extended Data Fig. 8 | See next page for caption.

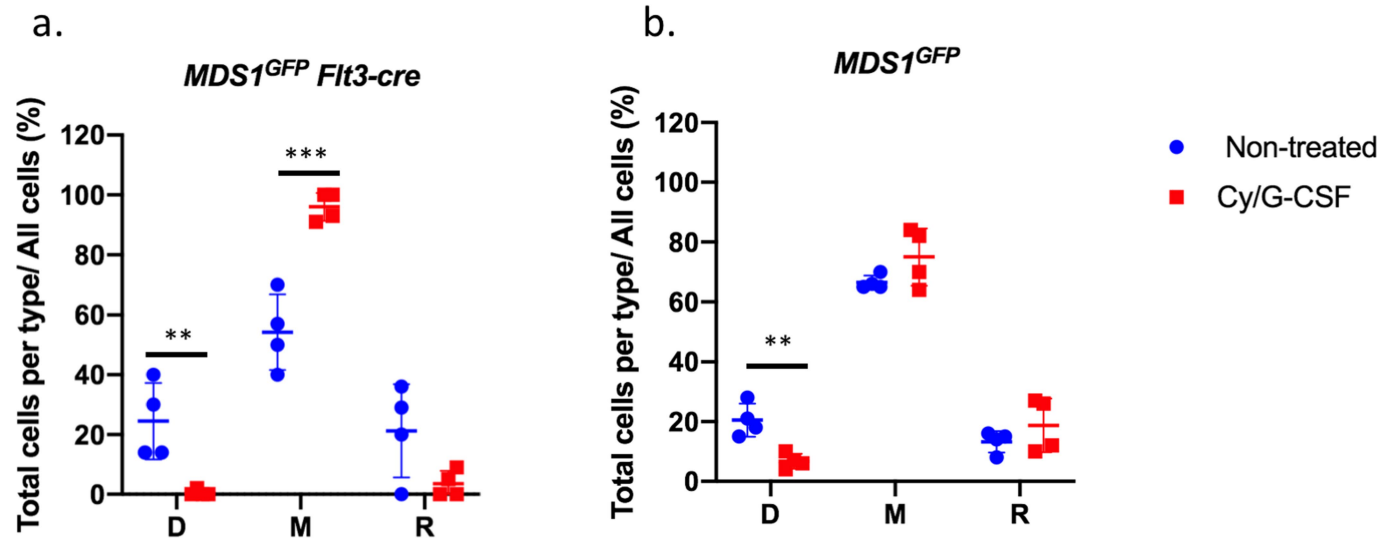
Extended Data Fig. 8 | Characterization of MFG-HSCs upon activation. a, Bone marrow analysis of HSPC (*Mds1^{GFP/+}*) PBS control (*n* = 1 mouse) and HSPC (*Mds1^{GFP/+}*) 5-FU-treated mice (*n* = 2 mice, value represents mean), 17 days after treatment. Data show marked expansion of HSPCs even after recovery of blood (Extended Data Fig. 1e). **b,** Graph showing in vivo motility measurements of MFG-HSCs at days 4 (*n* = 14 cells) and 20 (*n* = 13 cells) after 5-FU treatment. Red bar represents mean. Compare to untreated *Mds1^{GFP/+}Flt3^{Cre}* mice in Fig. 3a and Extended Data Fig. 7e. *P* was calculated using two-tailed Mann–Whitney test. **c,** Representative map of the locations of MFG-HSCs in the calvaria on day 20 after 5-FU treatment (*n* = 2 mice). Scale bar = 500 μ m. **d,** Generation of *Mds1^{CreER/+}Rosa26^{Confetti/+}* mice. **e,** Schematic illustration of Cy/GCSF treatment protocol for multicoloured *Mds1^{CreER/+}Rosa26^{Confetti/+}* labelling and activation. Low tamoxifen dosage (2 mg) was used to restrict recombination and expression of fluorescence in LT-HSCs that express higher levels of *Mds1*.

f, Detailed flow cytometry analysis of MPP3/4 cells, ST-HSCs and LT-HSCs with differential colour labelling upon treatment of *Mds1^{CreER/+}Rosa26^{Confetti/+}* mice shows labelling enriched in but not fully restricted to LT-HSCs. The experiment was performed once. **g,** 2D maps of the 3D locations of activated and labelled HSPCs in the fixed calvaria of control (left top, tamoxifen only, *n* = 2 mice) and induced mice (left bottom, tamoxifen + Cy/GCSF, *n* = 3 mice) along with maximum intensity projection (MIP) images (right top and bottom) of the *Mds1*-labelled cells (red, green, and blue). Scale bars for graphical map and MIP images, ~200 μ m and 50 μ m, respectively. **h,** Colour purity of cell clusters (original colours) compared to randomized colours (10,000 cycles) in three independent experiments (*n* = 3 mice). *P* values calculated using two-tailed unpaired *t*-test. Bar graphs with error bars represent mean and s.d., respectively.



Extended Data Fig. 9 | Validating bone cavity types using 2.3Col1-GFP (mature osteoblasts) and cathepsin K-activated fluorescent agent (osteoclasts). **a**, A montage of multiple z-stacks, displayed as the maximum intensity projection, showing double staining of bone marrow cavities in the calvarium. **b**, The same area as in **a**, showing the locations of 2.3Col1-GFP osteoblasts in areas of the old bone front that has not been eroded ($n = 3$ mice). **c**, Quantification of 2.3Col1-GFP pixels in D-type ($n = 10$ regions), M-type ($n = 16$ regions) and R-type cavities ($n = 18$ regions). Mean \pm s.d. **d**, A montage of multiple z-stacks, displayed as the maximum intensity projection, showing the double staining pattern (blue and red), 2.3Col1-GFP cells (green), osteoclasts (white), and bone marrow vasculature (purple). White arrows, osteoclast clusters. $n = 3$ mice. **e**, A zoomed-in region from **d** (box A), showing correlation

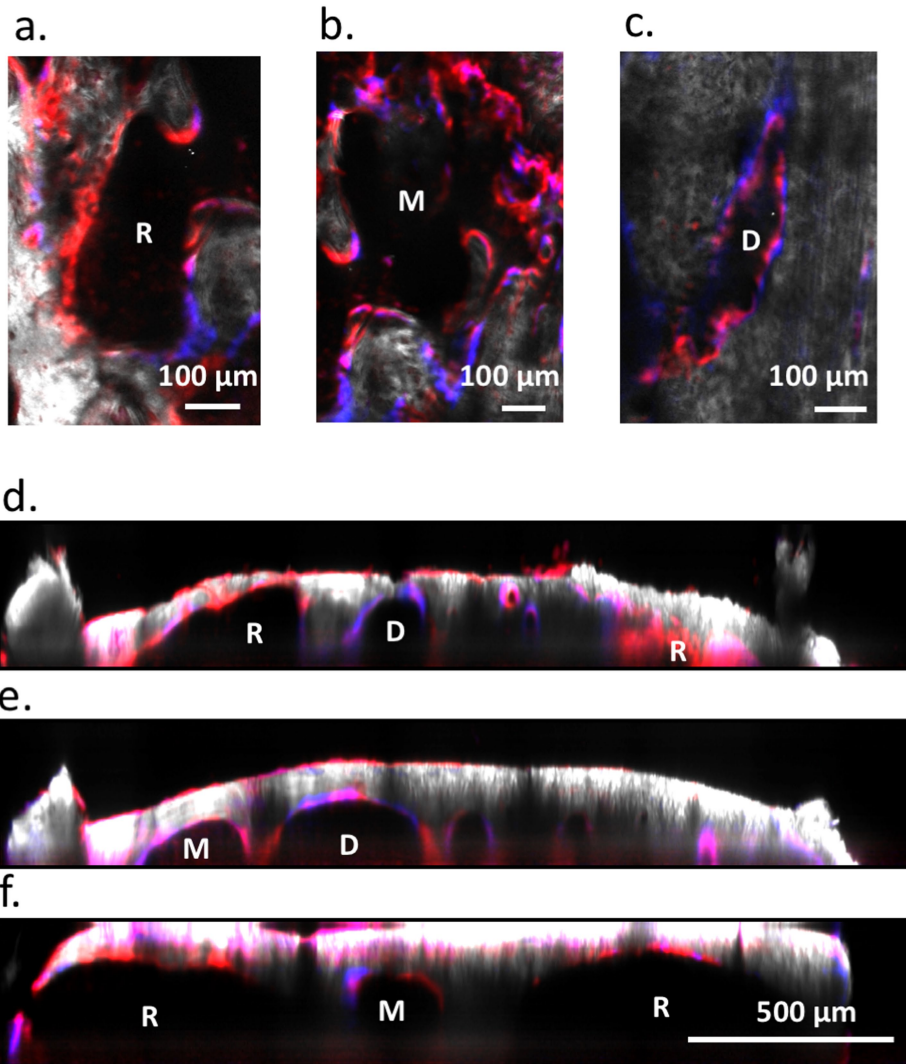
between 2.3Col1-GFP cells and the remaining dye 1 (blue) in a D-type cavity, and abundant cathepsin K⁺ osteoclasts in the R-type region where dye 1 was eroded. **f**, Examples of an M-type region from **d** (box B). In this region, dye 1 was eroded to some extent in spite of the presence of abundant 2.3Col1-GFP cells in the cavity. The corresponding cathepsin K panel shows the co-existence of several cathepsin K⁺ osteoclasts. **g**, Quantification of cathepsin K⁺ pixels in D-type ($n = 11$ regions), M-type ($n = 33$ regions), and R-type ($n = 10$ regions) cavities based on maximum intensity projection of montaged z-stacks. Compared to **c**, cathepsin K coverage shows a larger spread because it does not stain the cell body uniformly. Instead it frequently shows a punctate staining pattern, which is likely to represent lysosomes and endosomes. * $P < 0.0189$; ** $P = 0.0015$; **** $P < 0.0001$; two-sided Mann-Whitney test. Mean \pm s.d.



Extended Data Fig. 10 | Cell distribution in D-, M- and R-type cavities before and after Cy/G-CSF treatment. $n = 4$ mice per group. Graphs show the fractions of MDS or MFG cells distributed in D-, M- and R-type cavities at the steady state and after Cy/G-CSF treatment. The fraction is calculated by the total cells found in each cavity type divided by the total cells found in the calvaria of that mouse. **a**, The fractions of MFG cells increased in M-type cavities but decreased in D-type cavities after Cy/G-CSF treatment. Mean \pm s.d. Non-treated groups:

24.5 \pm 12.8, 54.3 \pm 12.6 and 21.3 \pm 15.6 in D-, M- and R-type cavities, respectively. Treated groups: 0.5 \pm 1.0, 96.0 \pm 4.7 and 3.5 \pm 4.4 in D-, M- and R-type cavities, respectively. ** $P = 0.0096$; *** $P = 0.0008$. **b**, The fractions of cells decreased in D-type cavities but remained the same in M- and R-type cavities. Mean \pm s.d. Non-treated groups: 20.5 \pm 5.6, 66.5 \pm 2.4 and 13.3 \pm 3.6 in D-, M- and R-type cavities, respectively. Treated groups: 6.8 \pm 2.5, 75.0 \pm 9.6 and 18.8 \pm 8.9 in D-, M- and R-type cavities, respectively. ** $P = 0.004$; unpaired, two-tailed t -test.

Dye 2 / Dye1



Extended Data Fig. 11 | Heterogeneous bone remodelling in bone marrow cavities of tibia metaphysis. A mechanically thinned metaphysis was imaged from the bone surface, labelled by sequential calcium staining. **a–c**, En face

views of D-, M- and R-type cavities from tibia metaphysis. **d–f**, x–z cross-section views from annotated white lines in Supplementary Video 15 show bone marrow cavities of varied remodelling stages similar to mouse calvaria.

Extended Data Table 1 | Activated MFG-HSCs (*Mds1*^{GFP/+} *Flt3*^{Cre} mice) are characterized by increased motility and various cellular interactions between GFP cells

	Total Cells	Paired Cell interactions	Mobilized cells
N=1	16	2	0
N=2	44	8	0
N=3	86	32	6
N=4	46	4	1
Total	192	46	7
Percentage		24.0%	3.6%

Table of observed MFG-HSC behaviours in Cy/GCSF-treated mice; n = 4.

Extended Data Table 2 | Summary table of findings from live imaging of native HSCs versus transplanted HSCs

Native HSCs (this work)	Transplanted HSCs
Adjacent to both endosteum and sinusoidal blood vessels	Adjacent to both endosteum and blood vessels (did not identify type) (Lo Celso C et al, Nature 2009)
Do not reside in regions with deepest hypoxia	Do not home in regions with deepest hypoxia (Spencer JA et al, Nature 2014)
Sessile; become motile after activation	Sessile; become motile after activation (Rashidi NM et al, Blood 2014)
Proliferate and form clusters after Cy/GCSF or 5-FU	Proliferate and form clusters after transplantation (Lo Celso C et al, Nature 2009)

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data collection methodology for calvaria imaging and full bone imaging is described in detail in the methods section. For full-bone immunofluorescence confocal imaging was performed using the default acquisition software of the Leica TCS SP8 confocal microscope as previously described (Coutu D.L. et al, Nature Methods 2018). Calavaria imaging data collection was performed as described in detail in the methods section and as previously described in Spencer J.A. et al, Nature 2014.

Data analysis

Data analysis methodology for calvaria imaging and full bone imaging is described in detail in the methods section. Custom code required for the analysis of full-bone immunofluorescence confocal imaging data was previously described (Coutu D.L. et al, Nature Methods 2018) and is available for download at <https://www.bsse.ethz.ch/csd/software/XiT.html>. In addition, Graph Pad Prism (version 7) and Imaris (8.3.1 and (9.1.2) were used for the corresponding data analysis and insertion of random dots for statistical comparison. For calvaria imaging data analysis Olympus FluoView software, Matlab, ImageJ scripts and Graph Pad Prism (version 6 or higher) were used. Several built-in plugins from image J were used, including contrast enhancement, 3D image J suite, background subtraction, global/ local thresholds, and 3D Euclidean distance measurements. For the single cell RNA sequencing analysis the code is available upon request. For single cell fluidigm experiments, data analysis and hierarchical clustering was performed using MultiExperiment Viewer (MeV) program. For the single cell RNA sequencing experiment, the data were processed using a previously published workflow and code available on <https://github.com/AllonKleinLab/SPRING>. Any additional python scripts used for graph plotting of the RNA sequencing data is described in detail in the methods section. For the synthesis and characterization of the phosphorescent oxyphor probe, MALDI-TOF was used to confirm the identity of the intermediate products and of the target probe molecule. To estimate the HSC frequency (MFG cells) in bone marrow we used the extreme limiting dilution analysis software available on <http://bioinf.wehi.edu.au/software/elda/>. GraphPad Prism and Excel were used to analyze and display all mouse characterization related data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw data from all in vivo experiments have been made available with the manuscript as source data in excel format or as supplementary information. The GEO accession code GSE115908 for the RNA seq data is publicly available.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculations were performed. The sample size was determined based on previous similar studies (Zhang Y et al, Blood 2011) and was adequate based on consistency of measured results in each group.
Data exclusions	Some data were excluded based on high variability and deviation from the mean in the % GFP positive cells present in Mds1-GFP/+ x Flt3-Cre mice. The 5 highest and 5 lowest values were excluded to ensure proper representation of the calculated % GFP positive. Exclusion criteria were not pre-established.
Replication	Experimental findings were reliably reproduced. In rare cases in which large variability was observed it is indicated with corresponding SD. To verify reproducibility of the findings the vast majority of experiments were repeated at least three independent times.
Randomization	Weaned mice from Mds1-GFP/+ x Flt3-Cre crosses, Mds1-CreER/+ x Rosa26-lox-stop-lox-Confetti crosses and Mds1-GFP/+ x C57/BL6 crosses were separated in male and female cages. Adult animals (2-6 months) of both sexes of corresponding genotypes were randomly selected and chosen for all experiments and used for BM isolation.
Blinding	No blinding was performed during data collection and analysis. For the study as we compare specific genotypes with or without treatment we have to pre-determine the genotype of each mouse followed by type of treatment performed and subsequent analysis. Thus, blinding during experiments and data acquisition is not possible in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials

All unique materials used are readily available from the authors.

Antibodies

Antibodies used

All antibodies were purchased from Biolegend, BD Biosciences, eBiosciences, Invitrogen, Cedarlane, Aves and R&D. The corresponding manufacturers per antibody are indicated in methods. All antibodies were used at 1:100 concentration unless otherwise stated.

Antibody Company Catalog number Concentration
 Flt3 PE-Cy5 eBioscience 15-1351-82 1:50
 c-kit APC eBioscience 17-1171-83 1:100
 c-kit AF700 eBioscience 56-1172-82 1:100
 Sca-1 PE-Cy7 Invitrogen 25-5981-82 1:100
 CD34 Biotin eBioscience 13-0341-85 1:33
 Fc R BV421 BioLegend 101331 1:50
 CD150 PE-Cy5 BioLegend 115912 1:100
 CD41 BV605 BioLegend 133921 1:100
 CD48 APC-Cy7 BD Pharmingen 561242 1:100
 CD45.2 V450 BD Horizon 560697 1:100
 B220 APC eBioscience 17-0452-83 1:100
 B220 Biotin eBioscience 13-0452-85 1:100
 CD19 APC-Cy7 eBioscience 47-0193-82 1:100
 CD19 Biotin Invitrogen 13-0193-85 1:100
 Mac1 APC-Cy7 eBioscience 47-0112-82 1:100
 CD4 APC Invitrogen 17-0041-83 1:100
 CD4 Biotin eBioscience 13-0041-85 1:100
 CD8a APC eBioscience 17-0081-83 1:100
 CD8a Biotin Invitrogen 13-0081-85 1:100
 Ly-6G AF700 BD Pharmingen 561236 1:100
 IgM eFluor 450 eBioscience 48-5890-82 1:100
 Gr-1 APC eBioscience 17-5931-82 1:100
 Gr-1 Biotin eBioscience 13-5931-85 1:100
 Ter119 PE-Cy5 eBioscience 15-5921-83 1:100
 Ter119 Biotin Invitrogen 13-5921-85 1:100
 Streptavidin APC eBioscience 17-4317-82 1:100
 Streptavidin APC-Cy7 eBioscience 47-4317-82 1:100
 Streptavidin PE-Cy7 eBioscience 25-4317-82 1:100
 Ki67 PE-Cy7 Biolegend 652426 1:100
 anti-GFP Aves Labs GFP-1020 1:50
 anti-CD117 R&D systems AF1356 1:50
 anti-collagen type I Cedarlane CL50151AP 1:50
 anti-CD105 eBioscience 14-1051-82 1:50
 Alexa Fluor 488 streptavidin conjugate Thermo Fischer Scientific S32354 1:50
 Alexa Fluor 555 Thermo Fischer Scientific A21432 1:50
 Alexa Fluor 633 Biotium 20137 1:50
 Alexa Fluor 680 Thermo Fischer Scientific A10043 1:50
 donkey anti-chicken biotin Jackson ImmunoResearch 703-065-155 1:50
 DAPI Thermo Fischer Scientific, D1306 1:2000

Validation

All antibodies used are well characterized and validated by providers. For all flow cytometry antibodies validation was performed in the mouse system using isotype control Abs as well as specific cell types that are known to be negative or positive for the corresponding Ab by the manufacturer. For all flow cytometry used Abs, validation of expression and fluorescence was performed using flow cytometry analysis by the manufacturer. All validation information for each Ab as well as previous publications that have used each Ab can be found on the manufacturer's website.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

All animals used in this study for BM and imaging analysis are of *Mus musculus* species, C57/BL6 background strain independent of genotype and 2-6 months of age. Both males and females were used for all experiments. Mds1-GFP/+, Mds1-GFP/+ Flt3-Cre, Mds1-CreER/+ Rosa26-CAG-lox-stop-lox-Confetti/+, Rosa26-CAG-lox-stop-lox-tdTomato/+ were generated as detailed described in methods or purchased from JAX and crossed with our generated strains. A detailed description of the procedure followed to generate Mds1-GFP/+, Mds1-GFP/+ Flt3-Cre and Mds1-CreER mouse lines is included in the methods section. For Tamoxifen induction and Cyclophosphamide/GCSF experiments the procedure as well as doses are detailed described in the methods section.

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve field-collected samples.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Bone marrow isolation was performed using crushing methodology, followed by red blood cell lysis. MACS beads and quadromACS (LS columns) were used for lineage depletion or c-kit enrichment. 40um filters were used to ensure single cell suspension prior FACS analysis. Antibody staining was performed for 45min, on ice, in PBS-2% FBS to ensure good staining and high viability of the cells. The detailed BM isolation protocol is included in the methods section.

Instrument

BD LSR II Flow cytometer was used for flow cytometry analysis, BD FACSAria II was used for cell sorting.

Software

BD FACSDIVA Software was used for data collection, FlowJo software (Tree Star) was used for data analysis.

Cell population abundance

Cells were sorted with Purity modes at 80-85% efficiency. Post sort fractions analyzed were at least 95% pure. Sorted samples were double sorted to ensure purity of the sorted populations. In experiments in which low cell numbers of sorted cells was used, cells were secondary sorted directly in plates to ensure accuracy of cell number.

Gating strategy

FSC/SSC preliminary gating was used to exclude debris (lower FSC) and to restrict the main bone marrow population including larger cells such as granulocytes that are found in higher SSC levels. Back-gating was used to ensure that all excluded populations were debris or dead cells (positive for DAPI) and that all included populations were part of the various positive antibody fractions. FSC-H vs. FSC-A was used to exclude doublet cells. SSC-H vs. SSC-A was used as an additional secondary doublet exclusion step. Dead cells were identified using DAPI staining in all experiments. Gating for negative/positive populations for all antibodies was performed using a negative control (no stain) followed by single color positive control for each antibody.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Neuronal programming by microbiota regulates intestinal physiology

<https://doi.org/10.1038/s41586-020-1975-8>

Received: 10 August 2019

Accepted: 9 December 2019

Published online: 5 February 2020

Yuuki Obata^{1*}, Álvaro Castaño¹, Stefan Boeing¹, Ana Carina Bon-Frauches¹, Candice Fung², Todd Fallesen¹, Mercedes Gomez de Agüero³, Bahtiyar Yilmaz³, Rita Lopes¹, Almaz Huseynova¹, Stuart Horswell¹, Muralidhara Rao Maradana¹, Werend Boesmans^{4,5}, Pieter Vanden Berghe², Andrew J. Murray⁶, Brigitta Stockinger^{1,7}, Andrew J. Macpherson^{3,7} & Vassilis Pachnis^{1,7*}

Neural control of the function of visceral organs is essential for homeostasis and health. Intestinal peristalsis is critical for digestive physiology and host defence, and is often dysregulated in gastrointestinal disorders¹. Luminal factors, such as diet and microbiota, regulate neurogenic programs of gut motility^{2–5}, but the underlying molecular mechanisms remain unclear. Here we show that the transcription factor aryl hydrocarbon receptor (AHR) functions as a biosensor in intestinal neural circuits, linking their functional output to the microbial environment of the gut lumen. Using nuclear RNA sequencing of mouse enteric neurons that represent distinct intestinal segments and microbiota states, we demonstrate that the intrinsic neural networks of the colon exhibit unique transcriptional profiles that are controlled by the combined effects of host genetic programs and microbial colonization. Microbiota-induced expression of AHR in neurons of the distal gastrointestinal tract enables these neurons to respond to the luminal environment and to induce expression of neuron-specific effector mechanisms. Neuron-specific deletion of *Ahr*, or constitutive overexpression of its negative feedback regulator CYP1A1, results in reduced peristaltic activity of the colon, similar to that observed in microbiota-depleted mice. Finally, expression of *Ahr* in the enteric neurons of mice treated with antibiotics partially restores intestinal motility. Together, our experiments identify AHR signalling in enteric neurons as a regulatory node that integrates the luminal environment with the physiological output of intestinal neural circuits to maintain gut homeostasis and health.

The enteric nervous system (ENS) encompasses the intrinsic neural networks of the gastrointestinal tract, which regulate most aspects of intestinal physiology (including peristalsis)^{6,7}. In addition to host-specific genetic programs, microbiota and diet have emerged as critical regulators of the physiology of gut tissue^{2,8} and changes in the microbial composition of the lumen often accompany gastrointestinal disorders⁴. Thus, depletion of the microbiota causes a reduced excitability of enteric neurons, changes in motility programs (such as the neurogenic colonic migrating motor complexes^{5,9,10}) and prolonged intestinal transit time (ITT)^{11,12}. However, conventionalization of adult germ-free mice reduces the deficit in ITT¹¹ and restores neuronal excitability¹³, which suggests that intestinal neural circuits are endowed with molecular mechanisms that monitor the state of the gut lumen and adjust neuronal activity and motility accordingly. Despite considerable recent progress² in describing the effects of the microbiota and diet on gastrointestinal physiology, the molecular mechanisms by which the

luminal environment regulates ENS activity and intestinal peristalsis remain unknown.

We hypothesized that molecular mechanisms that link the microbiota to intestinal motor behaviour are likely to be encoded by genetic programs that operate predominantly in neural circuits of the colon, the intestinal segment with the heaviest load of microorganisms¹⁴. We therefore used RNA sequencing to identify genes that are specifically upregulated in enteric neurons of the mouse colon in response to microbial colonization. Because our pilot experiments indicated that the current protocols for tissue dissociation and the recovery of intact ENS cells often resulted in considerable cellular damage and non-specific transcriptional changes, we developed a strategy that uses an adeno-associated virus (AAV) for labelling followed by the isolation and RNA sequencing of enteric neuron nuclei (nRNA-seq) that represent different intestinal segments and microbiota states (Fig. 1a, Extended Data Fig. 1a–l). First, we compared the transcriptional profiles of myenteric

¹The Francis Crick Institute, London, UK. ²Laboratory of Enteric Neuroscience (LENS), Translational Research in Gastrointestinal Disorders (TARGID), Department of Clinical and Experimental Medicine, University of Leuven, Leuven, Belgium. ³Maurice Muller Laboratories (DKF), Universitätsklinik für Viszerale Chirurgie und Medizin Inselspital, University of Bern, Bern, Switzerland. ⁴Biomedical Research Institute (BIOMED), Hasselt University, Hasselt, Belgium. ⁵Department of Pathology, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands. ⁶Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London, UK. ⁷These authors contributed equally: Brigitta Stockinger, Andrew J. Macpherson, Vassilis Pachnis. *e-mail: Yuuki.Obata@crick.ac.uk; Vassilis.Pachnis@crick.ac.uk

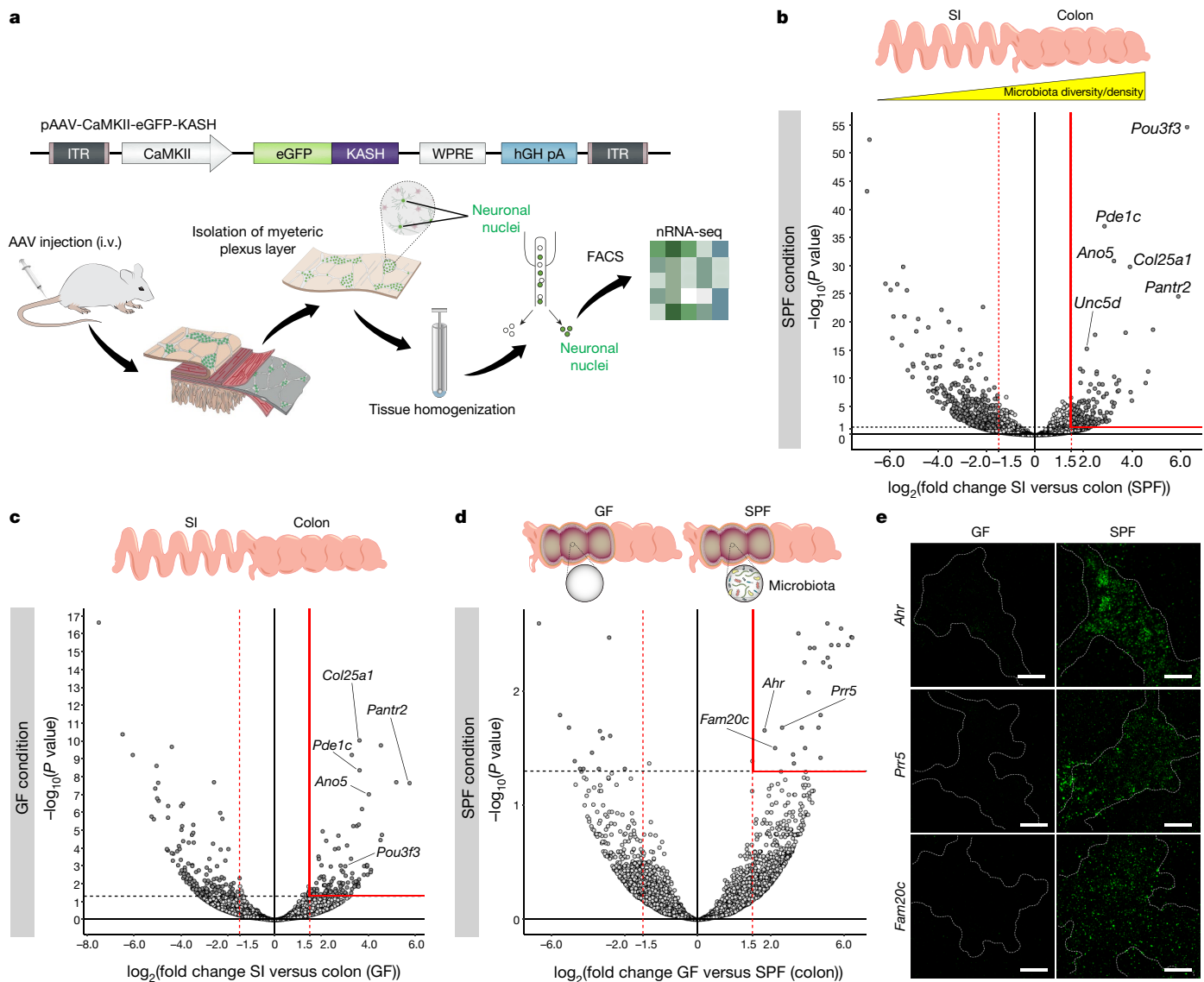


Fig. 1 | Programming of the enteric neuron transcriptome by microbiota.

a, Experimental design for AAV-mediated expression of nuclear-localized enhanced green fluorescent protein (eGFP) fused to the KASH nuclear membrane retention domain (eGFP-KASH)³⁰ in myenteric neurons and RNA sequencing of neuronal nuclei purified by fluorescence-activated cell sorting (FACS). The plasmid used to generate the AAV9-CaMKII-eGFP-KASH vector is shown at the top. i.v., intravenous. **b, c**, Volcano plots showing mean \log_2 -transformed fold change (x axis) and significance ($-\log_{10}(\text{adjusted } P \text{ value})$) of differentially expressed genes between myenteric neurons of the small intestine and the colon of SPF (**b**) and germ-free (GF) (**c**) mice. Differential gene-expression analysis was carried out using the DESeq2 R package. The default DESeq2 Wald test (two-sided) was used to determine P values for differential gene expression. Adjusted P values for differential gene expression were calculated with the DESeq2 default Benjamini-Hochberg method. \log_2 -transformed fold-change shrinkage was applied by having the betaPrior

parameter set to TRUE. Schematic of the small intestine and colon shown at the top. SPF CUEGs and germ-free CUEGs are within the area demarcated by red lines (\log_2 -transformed fold change = 1.5 < maximum; $P < 0.05$) in the plots of **b** and **c**, respectively. $n = 4$ SPF (Crick), 4 SPF (Bern) and 3 germ-free (Bern) independent nuclear isolates, each representing 3 mice. **d**, Volcano plot showing mean \log_2 -transformed fold change (x axis) and significance ($-\log_{10}(\text{adjusted } P \text{ value})$) of differentially expressed genes between colonic myenteric neurons from SPF and germ-free mice (Bern). Schematics of the colon from germ-free and SPF mice are shown at the top. Microbiota-dependent CUEGs are within the area demarcated by red lines (\log_2 -transformed fold change = 1.5 < maximum; $P < 0.05$). **e**, Representative images of myenteric ganglia (outlined by dotted line based on HuC/D immunostaining) from germ-free (left) and SPF (right) mice hybridized with RNAscope probes for *Ahr*, *Prr5* and *Fam20c*. Scale bars, 30 μm . Data represent two independent experiments.

neurons from the colon and small intestine of conventionally raised specific-pathogen-free (SPF) mice. To minimize potential variation in gene expression due to diet or other environmental factors, we used mice from two independent animal facilities (the Francis Crick Institute and University of Bern) to produce two datasets (the Crick and Bern datasets, respectively). Transcriptional profiles of neuronal nuclei clustered exclusively according to the intestinal segment of origin (Extended Data Fig. 1m), indicating that neurons derived from the small intestine or colon express distinct transcriptional programs.

Differential gene-expression analysis carried out on the combined Crick and Bern datasets identified 254 genes that were upregulated in enteric neurons of the colon (termed colon-upregulated ENS genes (CUEGs)) of SPF mice (hereafter, SPF CUEGs) (Fig. 1b, Supplementary Table 1). Among the top differentially expressed SPF CUEGs were genes that are implicated in neuronal development and function, such as *Pou3f3* (which encodes a transcription factor), *Ano5* (which encodes a chloride channel), *Pde1c* (which encodes a regulator of intracellular second messengers), *Unc5d* (which encodes a netrin receptor), *Col25a1*

(which encodes neuron-specific collagen) and *Pantr2* (which is a *Pou3f3*-adjacent non-coding transcript). To validate the nRNA-seq results, we used fluorescence RNA in situ hybridization (RNAscope) to analyse, at single-cell resolution, the relative expression of these genes in the myenteric plexus of the small intestine and colon. As expected, the hybridization signal was considerably stronger in enteric neurons of the colon relative to the small intestine (Extended Data Fig. 2a). These experiments demonstrate that the ostensibly homogeneous intrinsic neural networks along the mammalian intestine express segment-specific transcriptional programs.

To assess the contribution of microbiota to shaping the transcriptional landscape of neural circuits along the gut, we also compared the nuclear transcriptomes of enteric neurons from the small intestine and colon of germ-free mice. Using the same statistical criteria, we identified 122 CUEGs that are upregulated in colonic neurons of germ-free mice (hereafter, germ-free CUEGs) (Fig. 1c, Supplementary Table 2). The differential expression along the gut of many SPF CUEGs, including those we analysed by in situ hybridization, was maintained in germ-free mice (Extended Data Fig. 2b). These findings are consistent with parallel studies that demonstrated a similar expression of pan-neuronal and neuron-subtype markers in the myenteric plexus of SPF and germ-free mice (Extended Data Fig. 3a–f); this suggests that the transcriptional regionalization of the ENS along the mammalian gut is largely independent of microbial colonization, and that the effects of microbiota on ENS physiology are mediated by a small number of critical molecular pathways. To identify these pathways, we next compared directly the nuclear transcriptomes of colonic neurons from SPF and germ-free mice. We identified 25 genes (which we term microbiota-dependent CUEGs), the transcripts of which were more abundant in colonic neurons from SPF mice relative to germ-free mice (Fig. 1d, Supplementary Table 3). Several of the microbiota-dependent CUEGs (such as *Fam20c*) were absent from the list of SPF CUEGs, which suggests that they are likely to be expressed at comparable levels throughout the ENS but are under regulation by the microbiota specifically in colonic neurons. However, three fully annotated genes—*Ahr*, *Dand5* and *Prr5*—were also present in the list of SPF CUEGs (Supplementary Table 1), indicating that these genes are upregulated specifically in colonic neurons in response to microbiota colonization. RNAscope analysis confirmed the higher expression of *Ahr*, *Prr5* and *Fam20c* in colonic neurons from SPF mice relative to germ-free mice (Fig. 1e). Together, these studies reveal a previously unappreciated complexity of gene expression in the mammalian ENS and demonstrate that the transcriptional landscape of neural circuits along the gastrointestinal tract is shaped by the integrated effect of host-specific genetic programs and environmental factors such as the microbiota.

Next, we investigated the dataset of microbiota-dependent CUEGs as a potential source of regulatory and effector genes that link the microbial environment of the distal intestine with the functional output of colonic neural circuits. Initially, we focused on *Ahr* because it encodes a transcription factor with activity that is regulated by a broad range of microbial, dietary and endogenous metabolites (AHR ligands), and which functions as a biosensor that is critical for intestinal epithelial and immune-cell homeostasis^{15–17}. Upon ligand binding, cytosolic AHR translocates to the nucleus and induces expression (among others) of genes that encode cytochrome P450 (CYP1) enzymes, which metabolize AHR ligands and thereby terminate AHR signalling¹⁶. To provide evidence for a microbiota–AHR–neural-output axis in the gut, we immunostained the outer muscular layer (which includes the myenteric plexus) from the intestine of SPF and microbiota-manipulated mice for AHR. In the muscularis externa of the colon of SPF mice, AHR was expressed predominantly in myenteric neurons, which indicates that these neurons represent the main target of AHR ligands in this gut layer (Fig. 2a, Extended Data Fig. 4a–c). Almost all colonic myenteric neurons identified by the expression of the pan-neuronal markers HuC and HuD (designated hereafter as HuC/D) and subtype-specific (ChAT, nNOS,

calretinin, calbindin and NF-M) markers were positive for AHR (Fig. 2a, Extended Data Fig. 4d–g). Neurons exhibited either a cytoplasmic or a nuclear signal (Fig. 2a), suggesting differential activation of AHR across the population of enteric neurons. In contrast to the colon, myenteric neurons in the duodenum and the jejunum did not exhibit an AHR signal (Fig. 2b, Extended Data Fig. 4h), although a relatively weak signal was detected in the neurons of the distal ileum (Extended Data Fig. 4i). The lower expression of *Ahr* in myenteric neurons of the small intestine in comparison to the colon was also confirmed by quantification of the RNAscope in situ hybridization signal for AHR transcripts (Extended Data Fig. 4a, j–l). Together, these findings suggest that the expression of *Ahr* in enteric neurons is commensurate to the microbial load along the gut. In support of the microbiota-dependent expression of *Ahr*, myenteric neurons in the colon of germ-free mice and antibiotic-treated mice had reduced levels of AHR transcripts and a considerably weaker immunostaining signal, which was reinstated after colonization with the microbiota of SPF mice (Figs. 1e, 2c–e, Extended Data Fig. 4m–r). On the basis of these experiments, we suggest that microbiota-induced expression of *Ahr* in colonic neurons would enable lumen- or tissue-derived ligands to activate AHR signalling in colonic neural circuits, and thus contribute to the molecular profile and functional specialization of these circuits.

To identify potential target and effector genes of AHR signalling in the ENS, we next compared the nuclear transcriptomes of myenteric neurons from the colon of control mice and mice treated with the AHR ligand 3-methylcholanthrene (3MC)¹⁸. Among the 30 genes that showed the highest fold change with this treatment (which we term AHR-induced CUEGs) (Extended Data Fig. 5a) were *Ahr* and *Cyp1a1*, which are known transcriptional targets of *Ahr* in several types of non-neuronal cells¹⁵ and which have important roles in the feedback regulation of AHR signalling—either by repressing *Ahr*-dependent gene expression (*Ahr*) or by metabolizing AHR ligands (*Cyp1a1*)^{15,16}. AHR-dependent induction of *Cyp1a1* in enteric neurons was confirmed by quantification of the *Cyp1a1*-specific RNAscope signal in muscularis externa from the colon of control and 3MC-treated mice (Fig. 2f–h), and enhanced yellow fluorescent protein (eYFP) immunostaining of the myenteric plexus from 3MC-treated *Cyp1a1:cre;Rosa26eYFP* reporter mice¹⁶, in which activation of AHR signalling results in permanent expression of eYFP (Extended Data Fig. 5b–d). Querying the list of AHR-induced CUEGs for potential regulators of neuronal function downstream of the microbiota–AHR axis, we identified *Kcnj12* (Extended Data Fig. 5a), a gene that encodes the inwardly rectifying K⁺ channel, subfamily J member 12 (Kir2.2) that regulates the excitability of cardiac muscle and neuronal cells¹⁹. Previous studies have detected an inwardly rectifying current in mammalian enteric neurons²⁰, and the addition of a specific blocker (ML-133) of the current driven by the Kir2.x subfamily (Kir2.1, Kir2.2, Kir2.3 and Kir2.6)²¹ to live preparations of myenteric plexus altered the electrically evoked firing of enteric neurons (Extended Data Fig. 5e–g). *Kcnj12* was found among the top 30 microbiota-dependent CUEGs after marginal relaxation of the *P*-value criteria (*P* < 0.06) (Supplementary Table 3), which raises the possibility that this gene is regulated by microbiota and AHR signalling. In support of this idea, RNAscope experiments showed covariance of *Kcnj12* and *Cyp1a1* transcript levels in individual neurons in vivo after administration of 3MC (Fig. 2i–k) and—similar to *Ahr*—expression of *Kcnj12* was reduced in myenteric neurons from germ-free mice and antibiotic-treated mice (Fig. 2l–n, Extended Data Fig. 5h–j). We also observed a correlation between *Ahr* and *Kcnj12* expression in the enteric neurons of the colon of SPF mice under normal conditions (Extended Data Fig. 5k–m). To confirm that *Ahr* signalling regulates expression of *Kcnj12* in colonic neurons, we next used AAV-mediated gene transfer (Fig. 1a) to generate mice with enteric-neuron-specific deletion of *Ahr* (termed *Ahr*^{EN-KO} mice). An AAV9 vector expressing Cre recombinase under the control of the neuronal CaMKII promoter (AAV9-CaMKII-Cre), was administered to conditional *Ahr* mutant (*Ahr*^{fl/fl})²² and control

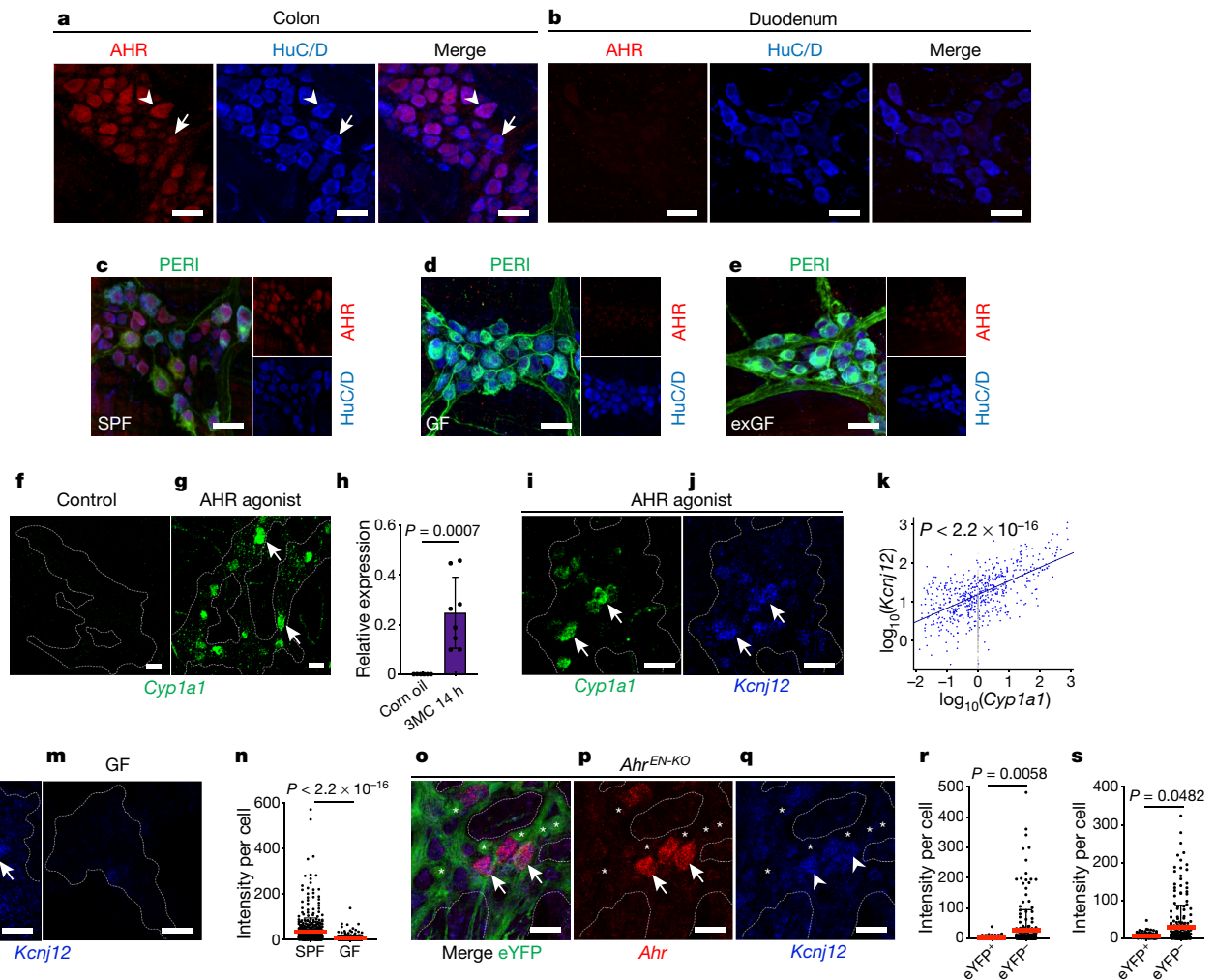


Fig. 2 | Microbiota- and ligand-dependent activation of AHR signalling.

a, b, AHR (red) and HuC/D (blue) immunostaining of colon (**a**) and duodenum (**b**) muscularis externa (12-week-old SPF mice). Note neurons with cytoplasmic (arrowhead) or nuclear (arrow) signals. $n = 12$ mice, 3 experiments. **c–e**, Immunostaining of SPF (**c**), germ-free (**d**) and conventionalized adult germ-free mice (exGF) (**e**) colonic myenteric ganglia with the pan-neuronal markers peripherin (PERI) (green), HuC/D (blue) and AHR (red). Small panels show AHR (top) or HuC/D (bottom). $n = 6$ mice for each condition, 2 experiments. **f, g**, Colonic myenteric ganglia from control (**f**) and 3MC-treated (**g**) mice hybridized with the *Cyp1a1* probe (green). Dotted line, borders of myenteric ganglia; arrows, positive neurons. $n = 4$ mice, 2 experiments. **h**, Quantitative PCR for *Cyp1a1* transcripts (mean \pm s.d.) in colonic muscularis externa from control and 3MC-treated mice. $n = 6$ control and 8 3MC-treated mice (two-sided non-parametric Mann–Whitney *U*-test). **i, j**, Colonic myenteric

ganglia from 3MC-treated mice hybridized with *Cyp1a1* (green) (**i**) and *Kcnj12* (blue) (**j**) probes. Arrows, neurons co-expressing *Cyp1a1* and *Kcnj12*. $n = 4$ mice, 2 experiments. **k**, Positive correlation in *Cyp1a1* and *Kcnj12* transcript level in myenteric neurons (*F*-test). $n = 827$ neurons, 4 mice. **l, m**, Colonic myenteric ganglia from SPF (**l**) and germ-free (**m**) mice hybridized with the *Kcnj12* probe. Arrows, positive neurons. $n = 6$ SPF mice and 6 germ-free mice, 2 experiments. **n**, Quantification of signal (mean \pm s.d.) in **l** and **m** (two-sided non-parametric Mann–Whitney *U*-test). $n = 510$ neurons from SPF mice and 298 neurons from germ-free mice. **o–q**, Myenteric ganglia of *Ahr*^{EN-KO} mice immunostained for eYFP (green) (**o**) and hybridized with the *Ahr* (red) (**p**) and *Kcnj12* (blue) (**q**) probes. **o**, Merge of signals in **p** and **q**. $n = 4$ mice, 2 experiments. **r, s**, RNA scope signal intensity (mean \pm s.d.) for *Ahr* (**r**) and *Kcnj12* (**s**) in eYFP⁺ (AHR⁺) and eYFP[−] (AHR[−]) neurons (two-sided non-parametric Mann–Whitney *U*-test). $n = 100$ eYFP⁺ and 223 eYFP[−] neurons, 4 mice. Scale bars, 30 μ m.

(*Ahr*^{+/+}) mice carrying the lineage reporter *Rosa26eYFP*, which enabled us to monitor the efficiency and cell-type specificity of *Ahr* deletion by GFP immunostaining (Extended Data Fig. 5n, o). Administration of AAV9-CaMKII-Cre to *Ahr*^{fl/fl}; *Rosa26eYFP* mice resulted in eYFP expression and ablation of *Ahr* from the majority of enteric neurons (Extended Data Fig. 5o, p). eYFP-expressing neurons—which, as expected, lack an AHR RNAscope signal (Fig. 2o, p)—showed significantly lower levels of *Kcnj12* transcripts relative to eYFP-negative neurons (Fig. 2o–s). Together, our experiments indicate that microbiota-dependent *Ahr* induction in enteric neurons enables the cell-autonomous activation of genes that encode previously known feedback regulators of AHR signalling as well as newly identified regulators of enteric neuronal excitability.

To determine whether AHR signalling in enteric neurons regulates intestinal physiology, we analysed the gut motility of *Ahr*^{EN-KO} mice. The

identification of AHR-deficient neurons for at least four weeks after AAV administration—along with the similar number, morphology and neurochemical properties of colonic neurons in control and *Ahr*^{EN-KO} mice (Extended Data Fig. 6a)—indicated that *Ahr* is dispensable for neuronal survival. Furthermore, no apparent deficit in the organization and cellular composition of the ENS was observed in constitutive *Ahr* mutant mice (*Ahr*^{−/−})²³ (Extended Data Fig. 6b). Despite the lack of discernible effects of *Ahr* deletion on the cellular organization of the ENS, the total ITT of *Ahr*^{EN-KO} mice was increased relative to two sets of control mice: AAV-injected wild-type mice (WT + AAV, which act as controls for the potential effects of AAV on ITT) and *Ahr*^{fl/fl} mice (cohousing with *Ahr*^{EN-KO} mice to minimize potential effects of microbiota) (Fig. 3a). WT + AAV and *Ahr*^{fl/fl} mice were indistinguishable from one another in terms of colon histology and motility (Extended Data Fig. 7a, b), but the increase

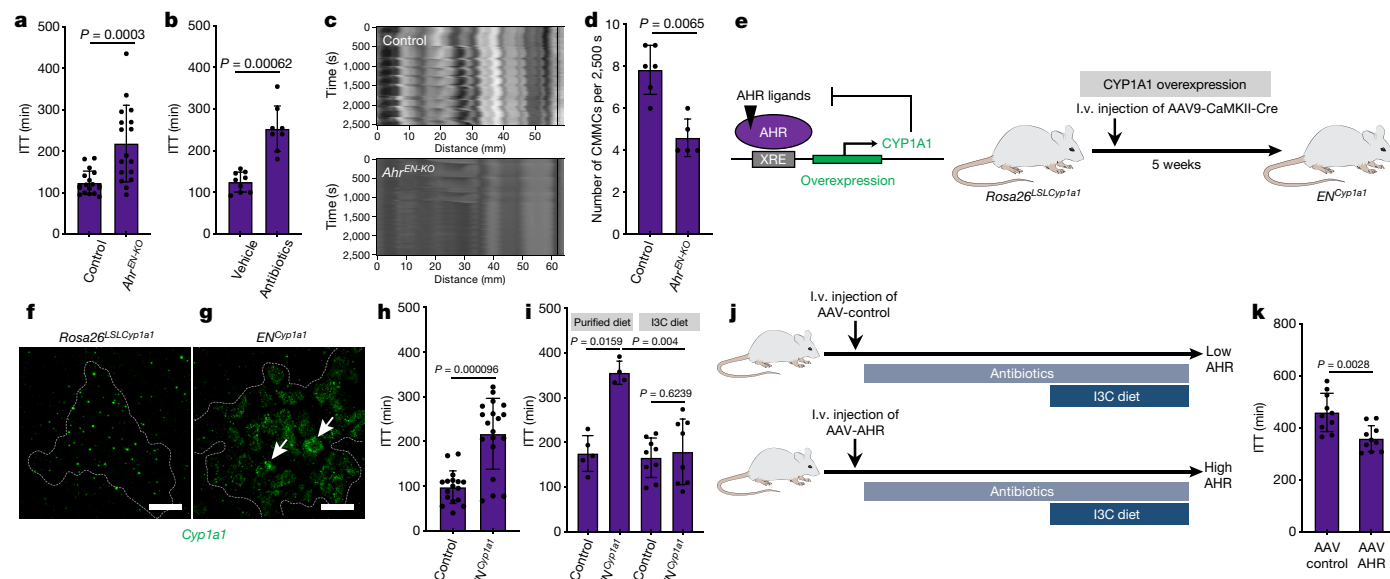


Fig. 3 | AHR signalling in enteric neurons regulates intestinal peristalsis. **a**, Quantification (mean \pm s.d.) of ITT in control and *Ahr*^{EN-KO} mice (two-sided non-parametric Mann–Whitney *U*-test). *n* = 18 control and 17 *Ahr*^{EN-KO} mice. **b**, Quantification (mean \pm s.d.) of ITT in vehicle- and antibiotic-treated mice (two-sided non-parametric Mann–Whitney *U*-test). *n* = 9 vehicle- and 8 antibiotic-treated mice. **c**, Representative ex vivo recorded colonic migrating motor complexes from control (top) and *Ahr*^{EN-KO} (bottom) mice. **d**, Quantification (mean \pm s.d.) of colonic migrating motor complexes (CMMCs) from control and *Ahr*^{EN-KO} mice (two-sided non-parametric Mann–Whitney *U*-test). *n* = 6 control and 5 *Ahr*^{EN-KO} mice. **e**, The negative feedback regulation of AHR signalling by CYP1A1 (left) is the basis for the experimental design to assess the role of neuron-specific *Cyp1a1* overexpression on intestinal motility (right). **f**, **g**, Myenteric ganglia from the colon of *Rosa26*^{LSL-Cyp1a1} (**f**) and *EN*^{Cyp1a1} (**g**) mice hybridized with the *Cyp1a1* RNAscope probe (green). Dotted line defines the borders of myenteric ganglia and arrows

indicate positive neurons. Data represent two independent experiments. Scale bars, 30 μ m. **h**, Quantification (mean \pm s.d.) of the effect of neuron-specific *Cyp1a1* overexpression on total ITT (two-sided non-parametric Mann–Whitney *U*-test). *n* = 17 control (WT + AAV and *R26*^{LSL-Cyp1a1}) and 19 *EN*^{Cyp1a1} mice. **i**, I3C-supplemented diet rescues the total ITT increase observed in *EN*^{Cyp1a1} mice (two-sided non-parametric Mann–Whitney *U*-test) (mean \pm s.d.). *n* = 5 control (purified diet), 4 *EN*^{Cyp1a1} (purified diet), 9 control (I3C diet) and 8 (*EN*^{Cyp1a1} (I3C-diet-fed) mice (female). **j**, Experimental design for expressing AHR in enteric neurons of microbiota-depleted mice. Wild-type SPF mice were injected with control (AAV-control) (top) or AHR-expressing (AAV-AHR) (bottom) AAV vectors and treated with antibiotics. All mice were fed with I3C-supplemented diet one week before ITT analysis. **k**, Quantification (mean \pm s.d.) of the effect of combinations of AAV-control and AAV-AHR vectors with antibiotic treatment on total ITT (two-sided non-parametric Mann–Whitney *U*-test). *n* = 10 mice per group.

in ITT after enteric-neuron-specific deletion of *Ahr* was comparable to that observed in microbiota-depleted SPF mice (Fig. 3a, b). To confirm that *Ahr* activity in enteric neurons regulates the physiological output of intestinal neural circuits independently of extrinsic gut innervation, we used ex vivo spatiotemporal mapping of colon preparations to record the ENS-dependent colonic migrating motor complexes¹⁰. The frequency and organization of colonic migrating motor complexes were reduced in *Ahr*^{EN-KO} mice (Fig. 3c, d), demonstrating that the cell-autonomous activity of *Ahr* in enteric neurons regulates the peristaltic activity of the colon. The CYP1A1-mediated clearance of natural AHR ligands¹⁶ predicts that constitutive upregulation of *Cyp1a1* in enteric neurons would phenocopy the effect of neuron-specific deletion of *Ahr* on intestinal motility, thus demonstrating the ligand-dependent activity of AHR in enteric neurons. To test this idea, we administered the AAV9-CaMKII-Cre vector to mice homozygous for the *Rosa26*^{LSL-Cyp1a1} allele¹⁶, resulting in constitutive overexpression of *Cyp1a1* specifically in enteric neurons (termed *EN*^{Cyp1a1} mice) (Fig. 3e–g). As expected, *EN*^{Cyp1a1} mice had an increased ITT (Fig. 3h) that is similar to that observed for *Ahr*^{EN-KO} mice (Fig. 3a), which indicates that dysregulation of AHR-ligand metabolism in enteric neurons disrupts intestinal motility. To examine further the potential role of AHR ligands in gut motility, we supplemented the diet of *EN*^{Cyp1a1} mice for four weeks with the AHR pro-ligand indole-3-carbinol (I3C), which generates the high-affinity ligand indolo[3,2-*b*]carbazole (ICZ)²⁴. Exposure to I3C diet rescued—to a large extent—the dysmotility in *EN*^{Cyp1a1} mice (Fig. 3i), which further demonstrates that the neuron-specific and ligand-dependent activation of AHR signalling regulates intestinal peristalsis. We suggest that, similar to ICZ, other natural ligands that originate in the gut lumen

and activate AHR in epithelial and immune cells in the gut wall^{16,25} are also capable of reaching nearby enteric neurons and their projections, modulating their transcriptional profile in an AHR-dependent manner.

Finally, to provide direct evidence that AHR signalling is implicated in the regulation of intestinal motility by microbiota, antibiotic-treated wild-type mice—which show reduced expression of *Ahr* in enteric neurons (Extended Data Fig. 4m–r) and a longer ITT (Fig. 3b)—were injected with AAV vectors expressing an AHR cassette under the control of the CaMKII promoter (*AAV9-CaMKII-Ahr*; AAV-AHR) or control vectors (AAV-control), and intestinal peristalsis was evaluated four weeks later. Because depletion of the microbiota is likely to reduce the amount of available AHR ligands^{26,27}, mice were also fed with I3C-supplemented diet for one week before the motility assay (Fig. 3j). As expected, antibiotic treatment of mice injected with AAV-control showed a marked increase in ITT, but injection of AAV-AHR resulted in a significant reduction of total transit time (Fig. 3k). However, the partial rescue that we observed suggested the presence of additional microbiota-dependent neuromodulators, such as serotonin (which is produced by enterochromaffin cells and modulates intestinal peristalsis)⁸. Together, our experiments demonstrate that AHR signalling in enteric neurons regulates the motor output of intestinal neural circuits.

In this study, we reveal regulatory mechanisms of enteric neurons that link the luminal microenvironment of the gut with ENS function. The systematic comparison of neuronal transcriptomes that represent distinct intestinal segments and microbiota states of mice enabled us to identify the transcription factor AHR as fulcrum of an ENS-specific surveillance pathway that regulates intestinal peristalsis in response to microbial colonization. Furthermore, the identification of genes

that regulate neuronal excitability (*Kcnj12*) as a class of AHR signalling targets provides a plausible molecular link between the dynamic micro-environment of the gut lumen and the functional output of intestinal neural circuits. Although the full spectrum and detailed molecular mechanisms that act downstream of neuronal AHR remain to be characterized, our experiments suggest that components of this pathway are implicated in the pathogenesis of intestinal motility disorders. A recent study has suggested a role for AHR in the biology of stool frequency, changes of which are one of the hallmarks of irritable bowel syndrome²⁸. Therefore, pharmacological or dietary interventions that modulate AHR activity in the cellular circuitry that controls intestinal peristalsis offer a realistic strategy for the management of conditions associated with gut dysmotility. In addition to the role of AHR in neurogenic motility, AHR-dependent transcriptional programs are also central to the barrier function of intestinal epithelial cells and the mucosal immune system^{25,29}. We suggest that, by transmitting environmental triggers within diverse cell types, AHR integrates the activity of functionally distinct intestinal tissues towards gut homeostasis and host defence.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1975-8>.

- Rolig, A. S. et al. The enteric nervous system promotes intestinal health by constraining microbiota composition. *PLoS Biol.* **15**, e2000689 (2017).
- Obata, Y. & Pachnis, V. The effect of microbiota and the immune system on the development and organization of the enteric nervous system. *Gastroenterology* **151**, 836–844 (2016).
- Dey, N. et al. Regulators of gut motility revealed by a gnotobiotic model of diet-microbiome interactions related to travel. *Cell* **163**, 95–107 (2015).
- De Palma, G. et al. Transplantation of fecal microbiota from patients with irritable bowel syndrome alters gut function and behavior in recipient mice. *Sci. Transl. Med.* **9**, eaaf6397 (2017).
- Hyland, N. P. & Cryan, J. F. Microbe–host interactions: influence of the gut microbiota on the enteric nervous system. *Dev. Biol.* **417**, 182–187 (2016).
- Yoo, B. B. & Mazmanian, S. K. The enteric network: interactions between the immune and nervous systems of the gut. *Immunity* **46**, 910–926 (2017).
- Furness, J. B. The enteric nervous system and neurogastroenterology. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 286–294 (2012).
- Yano, J. M. et al. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161**, 264–276 (2015).
- Vincent, A. D., Wang, X. Y., Parsons, S. P., Khan, W. I. & Huizinga, J. D. Abnormal absorptive colonic motor activity in germ-free mice is rectified by butyrate, an effect possibly mediated by mucosal serotonin. *Am. J. Physiol. Gastrointest. Liver Physiol.* **315**, G896–G907 (2018).
- Roberts, R. R., Murphy, J. F., Young, H. M. & Bornstein, J. C. Development of colonic motility in the neonatal mouse—studies using spatiotemporal maps. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292**, G930–G938 (2007).
- De Vadder, F. et al. Gut microbiota regulates maturation of the adult enteric nervous system via enteric serotonin networks. *Proc. Natl Acad. Sci. USA* **115**, 6458–6463 (2018).
- Ge, X. et al. Antibiotics-induced depletion of mice microbiota induces changes in host serotonin biosynthesis and intestinal motility. *J. Transl. Med.* **15**, 13 (2017).
- McVey Neufeld, K. A., Mao, Y. K., Bienenstock, J., Foster, J. A. & Kunze, W. A. The microbiome is essential for normal gut intrinsic primary afferent neuron excitability in the mouse. *Neurogastroenterol. Motil.* **25**, 183–e88 (2013).
- Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
- Stockinger, B., Di Meglio, P., Gialitakis, M. & Duarte, J. H. The aryl hydrocarbon receptor: multitasking in the immune system. *Annu. Rev. Immunol.* **32**, 403–432 (2014).
- Schiering, C. et al. Feedback control of AHR signalling regulates intestinal immunity. *Nature* **542**, 242–245 (2017).
- Gutiérrez-Vázquez, C. & Quintana, F. J. Regulation of the immune response by the aryl hydrocarbon receptor. *Immunity* **48**, 19–33 (2018).
- Stejskalova, L., Dvorak, Z. & Pavek, P. Endogenous and exogenous ligands of aryl hydrocarbon receptor: current state of art. *Curr. Drug Metab.* **12**, 198–212 (2011).
- Hibino, H. et al. Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol. Rev.* **90**, 291–366 (2010).
- Zholos, A. V., Baidan, L. V., Starodub, A. M. & Wood, J. D. Potassium channels of myenteric neurons in guinea-pig small intestine. *Neuroscience* **89**, 603–618 (1999).
- Wang, H. R. et al. Selective inhibition of the K_v2 family of inward rectifier potassium channels by a small molecule probe: the discovery, SAR, and pharmacological characterization of ML133. *ACS Chem. Biol.* **6**, 845–856 (2011).
- Walisser, J. A., Glover, E., Pande, K., Liss, A. L. & Bradfield, C. A. Aryl hydrocarbon receptor-dependent liver development and hepatotoxicity are mediated by different cell types. *Proc. Natl Acad. Sci. USA* **102**, 17858–17863 (2005).
- Schmidt, J. V., Su, G. H., Reddy, J. K., Simon, M. C. & Bradfield, C. A. Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proc. Natl Acad. Sci. USA* **93**, 6731–6736 (1996).
- Bjeldanes, L. F., Kim, J. Y., Grose, K. R., Bartholomew, J. C. & Bradfield, C. A. Aromatic hydrocarbon responsiveness-receptor agonists generated from indole-3-carbinol in vitro and in vivo: comparisons with 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Proc. Natl Acad. Sci. USA* **88**, 9543–9547 (1991).
- Metidji, A. et al. The environmental sensor AHR protects from inflammatory damage by maintaining intestinal stem cell homeostasis and barrier integrity. *Immunity* **49**, 353–362 (2018).
- Roager, H. M. & Licht, T. R. Microbial tryptophan catabolites in health and disease. *Nat. Commun.* **9**, 3294 (2018).
- Agus, A., Planchais, J. & Sokol, H. Gut microbiota regulation of tryptophan metabolism in health and disease. *Cell Host Microbe* **23**, 716–724 (2018).
- Jankipersadsing, S. A. et al. A GWAS meta-analysis suggests roles for xenobiotic metabolism and ion channel activity in the biology of stool frequency. *Gut* **66**, 756–758 (2017).
- Rothhammer, V. & Quintana, F. J. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. *Nat. Rev. Immunol.* **19**, 184–197 (2019).
- Wilhelmsen, K., Ketema, M., Truong, H. & Sonnenberg, A. KASH-domain proteins in nuclear migration, anchorage and other processes. *J. Cell Sci.* **119**, 5021–5029 (2006).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Mice

All mouse procedures at the Francis Crick Institute were carried out in accordance with the regulatory standards of the UK Home Office and approved by the local Animal Welfare and Ethics Review Body (AWERB). Procedures at the University of Bern were performed in accordance with Swiss Federal Regulations.

The following transgenic lines have previously been described: *Cyp1a1::cre* (ref. ³¹), *Ahr*^{-/-} (ref. ²³), *Ahr*^{R/R} (ref. ²²), *Rosa26*^{LSL-Cyp1a} (ref. ¹⁶), *Rosa26eYFP* (ref. ³²), *Wnt1::cre* (ref. ³³) and *Ai95D(RCL-GCaMP6f)-D* (ref. ³⁴). Details about the generation of the *ChAT-TVA-mCherry* mice have not been published; information about this line is available from V.P. upon request. Transgenic and wild-type C57BL/6 mice were bred and maintained in the SPF facility of the Francis Crick Institute. For most experiments, mice were on standard Crick diet. For the intestinal motility rescue experiment, mice were fed with purified diet supplemented with I3C (200 mg/kg) (Sniff Spezialdiäten), which generates metabolites representing physiological AHR ligands²⁴.

Wild-type and germ-free mice (C57BL/6) were bred and maintained in flexible-film isolators at the Clean Mouse Facility of the University of Bern (Switzerland). Germ-free status was monitored routinely by culture-based and other methods, and all mice were independently confirmed to be free of microorganisms³⁵. For bacterial colonization experiments, faecal contents of SPF mice were orally administered to wild-type germ-free mice and colonized mice were cohoused and maintained in the SPF facility of the University of Bern for four weeks before the analysis.

For depletion of the microbiota, wild-type (C57BL/6) mice were administered a 4-mM acetic acid solution containing 1 g/l ampicillin sodium, 0.5 g/l vancomycin hydrochloride, 1 g/l neomycin sulphate, 1 g/l metronidazole (all from Sigma-Aldrich) and 1% (v/w) artificial sweet flavour (Vimto) via drinking water for 18 days (Fig. 3k) or 30 days (Fig. 3b, Extended Data Figs. 4m–r, 5h–j).

For systemic administration of AAV vectors³⁶, 6-week-old mice were intravenously injected with AAV particles (dPCR GC titre > 1 × 10¹² GC) in 5% sucrose buffer (150 µl per mouse).

For activation of enteric neuronal AHR signalling, 8-week-old mice were injected intraperitoneally with AHR agonist 3MC³⁷ (26.5 mg/kg mouse), which is metabolized less efficiently in comparison to I3C. After treatment for 14 h (Fig. 2f–k, Extended Data Fig. 5a) or 5 days (Extended Data Fig. 5b–d), colonic muscular layers were isolated for nuclear purification or immunohistochemistry.

For our experiments we used male mice, unless specified otherwise in the Methods or relevant figure legends. All mice were between 8 and 16 weeks old.

Generation of AAV vectors

For the generation of the AAV vector expressing eGFP fused to the KASH nuclear membrane retention domain (eGFP–KASH)³⁰ under the control of the neuron-specific CaMKII promoter (AAV–CaMKII–eGFP–KASH), the DNA cassette encoding KASH-tagged eGFP (eGFP–KASH)³⁰ was amplified from the PX552 plasmid (Addgene, 60958) by Phusion High-Fidelity DNA polymerase. PCR was performed using the 5′-GCTATCGGATCCGCCACCATGGTG-3′ and 5′-GCTATCGAATTCCTAGG TGGGAGG-3′ primers, which enabled us to use the BamHI and EcoRI restriction sites to substitute eGFP–KASH for eGFP in plasmid pAAV–CaMKII–eGFP (Addgene, 50469), thus generating pAAV–CaMKII–eGFP–KASH (Fig. 1a). NEB stable competent *Escherichia coli* (New England Biolabs, C3040I) was transformed with pAAV–CaMKII–eGFP–KASH, and bacteria were grown on LB agar plates containing ampicillin

(50 mg/ml). The presence of inverted terminal repeats was confirmed by SmaI digestion. pAAV–CaMKII–eGFP–KASH has been deposited to Addgene (124882). Large-scale packaging of AAV–CaMKII–eGFP–KASH, AAV–CaMKII–Cre (Addgene, 105558) and AAV–CaMKII–eGFP (Addgene, 105541) vectors into serotype 9 capsid (AAV9) was carried out by Penn Vector Core. AAV9–CaMKII–AHR was generated by Vector Biolabs.

Histopathological analysis

Intestinal segments were collected from mice infected with AAV9 vectors, fixed with 4% paraformaldehyde (PFA) in PBS overnight at 4 °C and embedded in paraffin. Paraffin was removed from sections with xylene, rehydrated with ethanol and stained with either haematoxylin and eosin or Alcian blue–PAS haematoxylin.

Immunohistochemistry

For immunostaining, the intestine was flushed free of luminal contents, cut along the mesenteric border following removal of adipose tissues on serosa and fixed with 4% PFA overnight at 4 °C. For the preparation of the submucosal plexus layer, a 1-ml pipette was inserted into the lumen to fully extend the smooth muscle layer containing myenteric plexus, which was removed from the mucosal compartment using cotton buds as previously described³⁸. The submucosal plexus layer was treated with 30 mM EDTA in PBS for 30 min on ice to remove the epithelial layer, stretched on Sylgard coated Petri dish and fixed with 4% PFA for 3 h at 4 °C. The fixed tissues were then rinsed with PBS for three times at room temperature. Gut tissues were then permeabilized and preblocked with 10% normal donkey serum (NDS) and 1% Triton X-100 in PBS for 1 h at room temperature and incubated with primary antibodies (listed in the Reporting Summary) in the same buffer for 48 h (at 4 °C). Tissue was then incubated with secondary antibodies (listed in the Reporting Summary) in 10% NDS and 1% Triton X-100 for 12 h at room temperature. DAPI (Molecular Probes, d3571) was used to counterstain the nucleus. Samples were washed with PBS before mounting with VECTASHIELD (Vector Laboratories).

Image processing

Immunostained gut preparations were examined with Olympus FV3000-Invert (SW312-CB1) confocal laser scanning microscope and FV31S-SW software (Olympus) using standard excitation and emission filters for visualizing DAPI, Alexa Fluor 488, Alexa Fluor 568 and Alexa Fluor 647. All images were processed with Adobe Photoshop CS 8.0 (Adobe Systems) while analyses were performed using the image-processing package Fiji and ImageJ (W. Rasband, NIH).

Fluorescence in situ hybridization

Fluorescence in situ hybridization on the myenteric plexus was carried out using the Advanced Cell Diagnostics RNAscope Fluorescent Multiplex Kit (ACD, 320850) according to manufacturer's specification. In brief, the muscularis externa layer of the gut was dehydrated by serial ethanol treatments and treated with RNAscope Protease III for 30 min (for small intestine) or 45 min (for colon) at room temperature. Tissue was then incubated overnight (at 40 °C) with fluorescent probes, 3-plex positive control probe, 3-plex negative control probe or customized probes. Following hybridization, tissue was washed twice with wash buffer and then subjected to sequential hybridization with pre-amplifier, amplifier DNA (amp1-FL, amp 2-FL and amp 3-FL) and fluorophore (amp 4 alt A-FL) at 40 °C for 15–30 min for each step. After hybridization, tissues were counterstained for the pan-neuronal marker HuC/D) and mounted on Superfrost Plus Adhesion Microscope Slides (ThermoFisher Scientific, 10149870) using VectaMount Permanent Mounting Medium (ACD, 321584). RNAscope probes (Advanced Cell Diagnostics) used in this study include Ret-C1 (431791), Ahr-C1 (452091), Cyp1a1-C1 (464611), Pou3f3-C1 (441521), Pde1c-C1 (489011), Ano5-C1 (557141), Pantr2-C1 (483721), Prr5-C1 (557121), Fam20c-C1 (453351), Unc5d-C2 (480461), Col25a1-C3 (538511) and Kcnj12-C3 (525171).

Quantification of RNAscope intensity on enteric neurons

Fluorescent signals of RNAscope were colocalized with HuC/D⁺ enteric neurons using an automated pipeline in CellProfiler³⁹. In brief, the Alexa Fluor 405 channel (HuC/D) was background-corrected using a baseline subtraction of 10% of the maximum pixel intensity. The background-subtracted image was then segmented and individual neurons were detected using the IdentifyPrimaryObjects module and clumped objects were separated by shape. RNAscope spots (2–30 pixels) were identified separately for each gene and were processed further only if colocalized to HuC/D⁺ neurons. Individual neurons were tracked through the z-stack by measured overlap. The RNAscope signal intensity of individual neurons was integrated through the image z-stack in slices in which the cells were identified.

Quantitative PCR

Total RNA was isolated from the colonic muscular layer using Trizol LS reagent and the PureLink RNA Micro Kit (Invitrogen, 12183016) according to the manufacturer's specifications, and was subjected to reverse transcription using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 4368814). Quantitative PCR was performed with complementary DNA (cDNA) using Taqman fast universal 2× PCR Master Mix (Applied Biosystems) and Taqman probes (Applied Biosystems) for *Actb* (Mm02619580_g1) and *Cyp11a1* (Mm00487218_m1). *C_t* values obtained were normalized to *Actb*.

Purification of neuronal nuclei from the myenteric plexus

For the isolation of neuronal nuclei from the myenteric plexus, mice were injected intravenously with AAV9-CaMKII-eGFP-KASH. Five weeks after intravenous injection of the AAV, the longitudinal smooth muscle layer and associated myenteric plexus were peeled off the wall of the small intestine and colon and subjected to Dounce homogenization in lysis buffer (250 mM sucrose, 25 mM KCl, 5 mM MgCl₂, 10 mM Tris buffer with pH 8.0, 1 mM DTT) containing 0.1% Triton X-100, cOmplete EDTA-free protease inhibitor (Sigma-Aldrich) and DAPI. After filtering the homogenate to remove large debris, samples were centrifuged at 1,000g for 10 min at 4 °C to obtain a pellet containing muscularis externa nuclei. For flow cytometric analysis, doublet discrimination gating was applied to exclude aggregated nuclei, and intact nuclei were determined by subsequent gating on the area and height of DAPI intensity. Both eGFP⁺ and eGFP[−] nuclear populations were collected directly into a 1.5-ml tube containing Trizol LS reagent (Invitrogen) using, at the Francis Crick Institute, the Aria Fusion cell sorter (BD Biosciences) and, at the University of Bern, the Aria III (BD Biosciences). The obtained FCS data were further analysed using FlowJo software version 10.5.3.

The list of SPF CUEGs was generated by combining nRNA-seq data from mice housed in both facilities. Labelling and isolation of neuronal nuclei in the two facilities was carried out using exactly the same protocol and reagents, except for the use of two different FACS sorters (Aria Fusion at the Francis Crick Institute and Aria III in Bern). RNA extraction and RNA sequencing for all samples was done at the Francis Crick Institute. The list of germ-free CUEGs was generated with mice sourced exclusively from the Bern germ-free facility. The identification of microbiota-dependent CUEGs was done by comparing the transcriptome of colonic neurons from SPF and germ-free mice from the Bern facility. Finally, the transcriptomic experiment that generated the list of AHR-induced CUEGs was carried out using exclusively Crick mice.

RNA sequencing and bioinformatic analysis

Extraction of nuclear RNA and nuclear RNA sequencing were carried out at the Francis Crick Institute. Nuclear RNA was isolated using PureLink RNA Micro Kit (Invitrogen, 12183016) according to the manufacturer's instructions. Double-stranded full-length cDNA was generated using the Ovation RNA-Seq System V2 (NuGen Technologies). Following quantification on a Qubit 3.0 fluorometer (Thermo Fisher Scientific),

cDNA was fragmented to 200 bp by acoustic shearing using Covaris E220 instrument (Covaris) at standard settings. The fragmented cDNA was then normalized to 100 ng, which was used for sequencing library preparation using the Ovation Ultralow System V21-96 protocol (NuGen Technologies). A total of 7 PCR cycles was used for library amplification. The quality and quantity of the final libraries were assessed with TapeStation D1000 Assay (Agilent Technologies). The libraries were then normalized to 2.5 nM, pooled and loaded onto a HiSeq 4000 (Illumina) to generate 75-bp single-end reads. For the transcriptomic comparison of colon to small-intestine myenteric neurons from SPF mice (Fig. 1b) we used eight nuclear isolates (four from Crick, and four from Bern, mice), each representing three mice. All mice in the Crick samples were male. Two of the Bern samples were generated from male and two from female mice. Male and female samples were indistinguishable by principal component analysis. The transcriptomic comparison of colon to small-intestine myenteric neurons from germ-free mice (Fig. 1c) were generated using three nuclear isolates (all Bern mice), each representing three mice.

For the bioinformatics analysis, the 'Trim Galore!' utility version 0.4.2 was used to remove sequencing adaptors and to quality trim individual reads with the *q*-parameter set to 20. The sequencing reads were then aligned to the mouse genome and transcriptome (Ensembl GRCm38 release-86) using RSEM version 1.3.0 in conjunction with the STAR aligner version 2.5.2. Sequencing quality of individual samples was assessed using FASTQC version 0.11.5 and RNA-SeQC version 1.1.8. Differential gene expression was determined using the R Bioconductor package DESeq2 version 1.14.1.

ITT assay

The total ITT was measured as previously described³⁸. Mice were placed individually in bedding-free cages with the diet and HydroGel (Clear H₂O). For all experiments, 250 µl of 6% (w/v) carmine red dye (Sigma-Aldrich) in 0.5% (w/v) methylcellulose (Sigma-Aldrich) was orally administered to each mouse at 09:00. The time period from gavage until the emergence of the first red-colour pellet was recorded as total ITT.

Live video imaging and spatiotemporal mapping of colonic motility

Ex vivo video imaging and analysis of colonic motility was performed as previously described³⁸. In brief, the entire colon was carefully isolated and loosely pinned in an organ bath chamber continuously superfused (flow rate of 4 ml/min) with oxygenated (95% O₂ and 5% CO₂) Krebs solution (in mM, 120.9 NaCl, 5.9 KCl, 1.2 MgCl₂, 2.5 CaCl₂, 1.2 NaH₂PO₄, 14.4 NaHCO₃, 11.5 glucose) kept at 37 °C. Following equilibration of the colon for 30 min, movies of colonic motility were captured (2.5-Hz frame rate) with a QICAM-Fast camera using QCapture Pro 6.0 software (Q-Imaging). Images were read into an Igor Pro (WaveMetrics) and analysed using custom-written algorithms. The edges of the bowel were determined, and the width computed and mapped over time. From the generated spatiotemporal maps, the frequency of propagating contractions was determined.

Ca²⁺ imaging of colonic myenteric plexus

For ex vivo Ca²⁺ imaging experiments, the large intestine of adult *Wnt1::cre;Rosa26-GCaMP6f* mice^{33,34} was isolated and pinned flat in a Sylgard-lined dish filled with Krebs solution, bubbled with 95% O₂ and 5% CO₂ at room temperature. The mucosal, submucosal and longitudinal muscle layers were carefully removed to obtain a circular muscle with adherent myenteric-plexus preparation, which was mounted over a small inox ring, immobilized by a matched rubber O-ring⁴⁰. Myenteric-plexus preparations were placed in a recording chamber mounted on an upright Zeiss Axio Examiner.Z1 microscope equipped with a Poly V xenon monochromator (TILL Photonics) and water dipping lens (20×, NA 1.0, Zeiss). GCaMP6f was excited at 475 nm and images were recorded at 525/50 nm (at 2 Hz) on a Sensicam-QE CCD camera

(PCO) using TillVisION (TILL Photonics). Myenteric plexus preparations were constantly superfused with carbogenated Krebs solution at room temperature containing 1 μ M nifedipine (Sigma-Aldrich) via a local gravity-fed (\pm 1 ml/min) perfusion pipette. Myenteric ganglia were stimulated electrically using single and a train (20 Hz, 2 s) of pulses (300 μ s) transmitted from a Grass stimulation unit via a focal electrode (50- μ m diameter tungsten wire) placed on an interganglionic connective leading to the selected myenteric ganglia within the field of view.

Analysis was performed with custom-written routines in Igor Pro (Wavemetrics)⁴⁰. Regions of interest were drawn, after which the average Ca^{2+} signal intensity was calculated, normalized to the initial GCaMP6f values and reported as F/F_0 . Cells were considered as responders when the GCaMP6f signal rose above baseline plus 3 \times the intrinsic noise (standard deviation) during the recording. The Ca^{2+} transient amplitudes were measured as the maximum increase in $[\text{Ca}^{2+}]_i$ above baseline (maximum F_i/F_0).

Statistical analysis

Statistical comparisons between samples were performed in GraphPad Prism software using Student's *t*-test. When variances were not homogeneous, the data were analysed by the non-parametric Mann-Whitney *U*-test. For correlation analysis of RNAscope intensity (Fig. 2k, Extended Data Fig. 5m), data were filtered to remove pairs for which either observation was zero and data were then \log_{10} -transformed. The correlation coefficient was calculated by fitting a standard linear regression model to this \log_{10} -transformed data and quality of the overall model was assessed with an *F*-test. Data excluded from the analysis were overlaid on the scatter plot by \log_{10} -transforming the non-zero observation and retaining the zero value. The analysis was performed using R 3.3.1.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All RNA-seq data are available at Gene Expression Omnibus (GEO) under accession number GSE140293. Source Data for Figs. 2, 3 and Extended Data Fig. 1–5, 7 are provided with the paper. All datasets analysed during the current study are presented in this manuscript, or are available from the corresponding authors upon reasonable request.

Code availability

The source code and installation instructions for colonic migrating motor complex evaluation and Ca^{2+} imaging can be found at <https://doi.org/10.7554/eLife.42914.039> (Ca^{2+} imaging analysis source code) and <https://doi.org/10.7554/eLife.42914.040> (installation instructions and user guide). For more information, please contact pieter.vandenbergh@kuleuven.be. The code related to the RNAscope signal

quantification is available at GitHub (<https://github.com/FrancisCrick-Institute/Pachnis-lab/tree/master/Neuronal-programming-Nature>).

31. Henderson, C. J. et al. Application of a novel regulatable Cre recombinase system to define the role of liver and gut metabolism in drug oral bioavailability. *Biochem. J.* **465**, 479–488 (2015).
32. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
33. Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. & McMahon, A. P. Modification of gene activity in mouse embryos *in utero* by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8**, 1323–1326 (1998).
34. Madisen, L. et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
35. Gomez de Agüero, M. et al. The maternal microbiota drives early postnatal innate immune development. *Science* **351**, 1296–1302 (2016).
36. Gombash, S. E. et al. Intravenous AAV9 efficiently transduces myenteric neurons in neonate and juvenile mice. *Front. Mol. Neurosci.* **7**, 81 (2014).
37. Jiang, W. et al. Persistent induction of cytochrome P450 (CYP)1A enzymes by 3-methylcholanthrene *in vivo* in mice is mediated by sustained transcriptional activation of the corresponding promoters. *Biochem. Biophys. Res. Commun.* **390**, 1419–1424 (2009).
38. Sasselli, V. et al. Planar cell polarity genes control the connectivity of enteric neurons. *J. Clin. Invest.* **123**, 1763–1772 (2013).
39. Kamentsky, L. et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).
40. Li, Z. et al. Regional complexity in enteric neuron wiring reflects diversity of motility patterns in the mouse large intestine. *eLife* **8**, e42914 (2019).
41. Rao, M. et al. Enteric glia express proteolipid protein 1 and are a transcriptionally unique population of glia in the mammalian nervous system. *Glia* **63**, 2040–2057 (2015).
42. Gabanyi, I. et al. Neuro-immune interactions drive tissue programming in intestinal macrophages. *Cell* **164**, 378–391 (2016).

Acknowledgements We thank the Crick Science Technology Platforms, the University of Bern FACS Lab and the Bern Clean Mouse Facility for expert support; R. Lasrado and S.-H. Chng for assistance with tissue dissection; M. Shapiro for bioinformatic input; C. Schiering for useful advice; all members of the Pachnis laboratory for insightful comments on the manuscript and discussions; and M. D'Amato for insightful comments on the manuscript. Y.O. was supported by an EMBO long-term fellowship (ALTF 1214-2015), travel grants from Boehringer Ingelheim Fonds and the Society for Mucosal Immunology (SMI); he is currently supported by an HFSP postdoctoral fellowship (LT000176/2016). This work was supported by the Medical Research Council (MRC) and The Francis Crick Institute (which receives funding from the MRC, Cancer Research UK and the Wellcome Trust). V.P. was also funded by BBSRC (BB/L022974) and the Wellcome Trust (212300/Z/18/Z).

Author contributions Y.O. and V.P. conceived the study and together with B.S. and A. J. Macpherson designed the experiments. Á.C., A.C.B.-F., C.F., M.G.d.A., B.Y., M.R.M., W.B. and B.Y. helped with the experiments. Á.C. performed the RNAscope *in situ* hybridization experiments; T.F. helped with the quantification of RNAscope data; M.G.d.A. and B.Y. helped to organize experiments with germ-free and exGF mice; A.C.B.-F., W.B. and V.B. provided help with the spatiotemporal mapping experiments and the analysis. C.F. carried out and analysed the Ca^{2+} imaging experiments with help from W.B. and P.V.B. A.H. prepared the cDNA library for the bulk nRNA-seq. S.B. performed bioinformatics analysis. S.H. performed statistical analysis. R.L. generated *ChAT-TVA-mCherry* mice. Y.O. generated the AAV-CaMKII-eGFP-KASH construct with help from A. J. Murray. V.P. and Y.O. wrote the manuscript with help from B.S. and A. J. Macpherson, and contributions from all authors.

Competing interests The authors declare no competing interests.

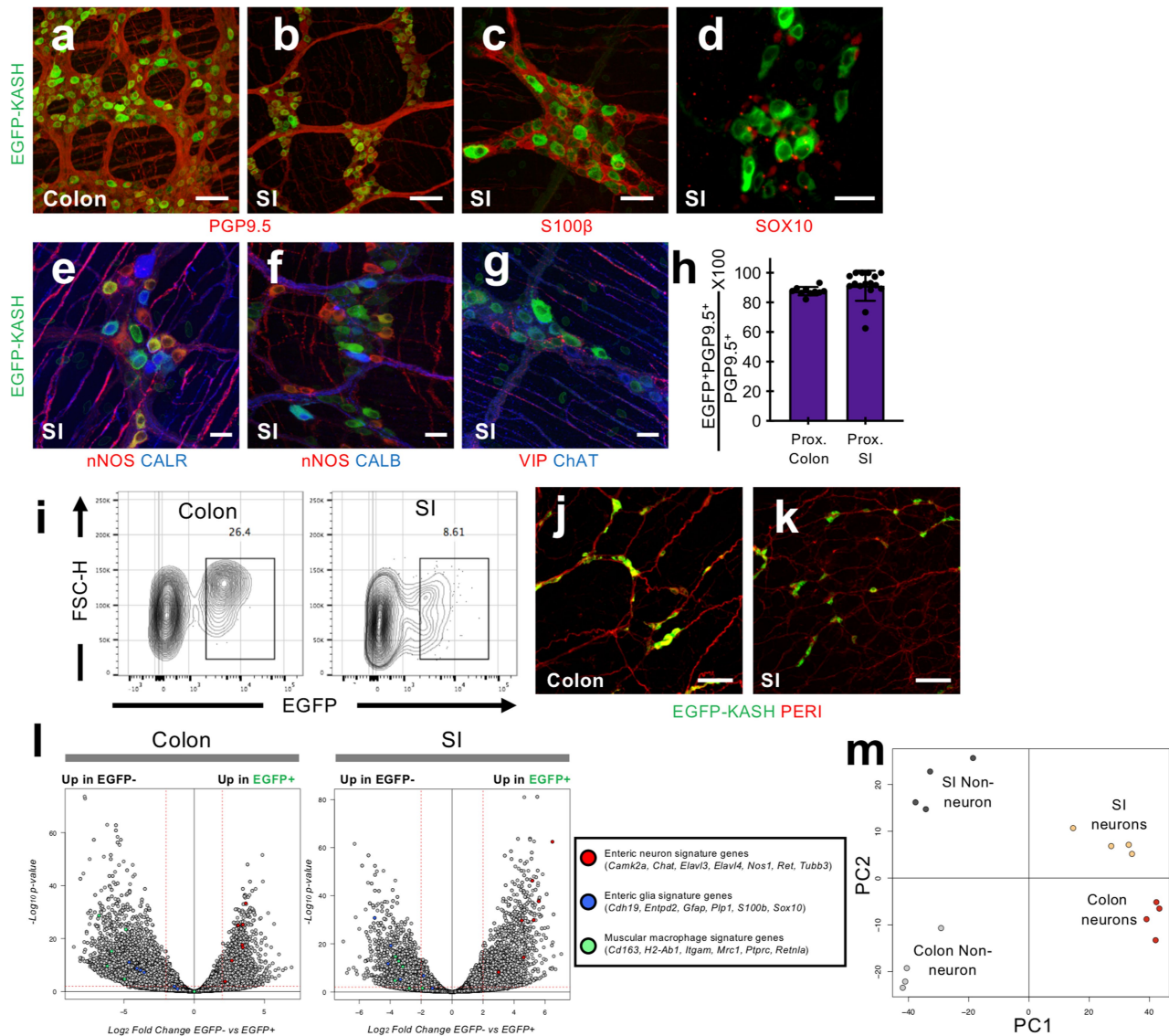
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1975-8>.

Correspondence and requests for materials should be addressed to Y.O. or V.P.

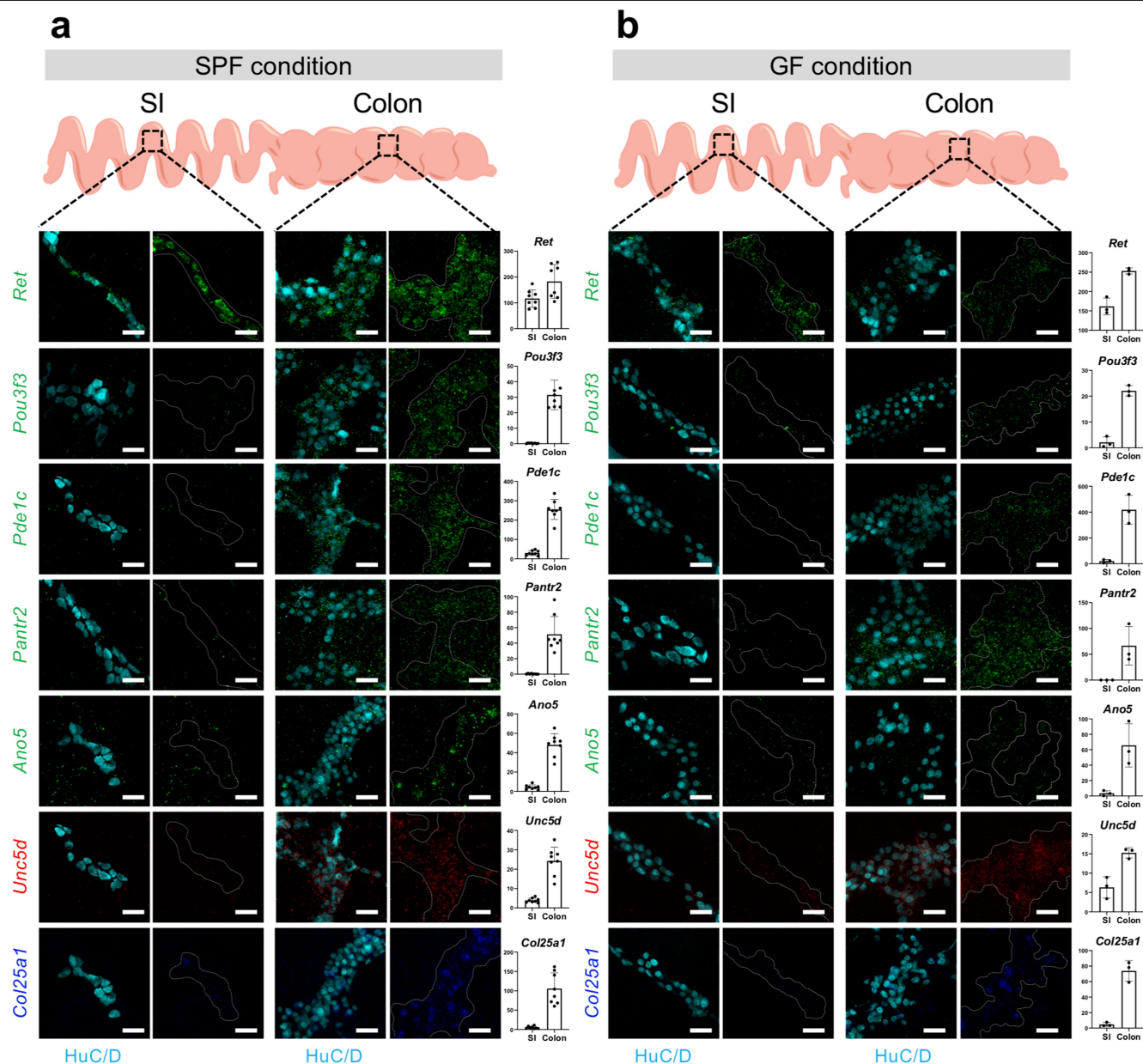
Peer review information *Nature* thanks John Cryan, Michael D. Gershon, Sven Pettersson and Harry Sokol for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



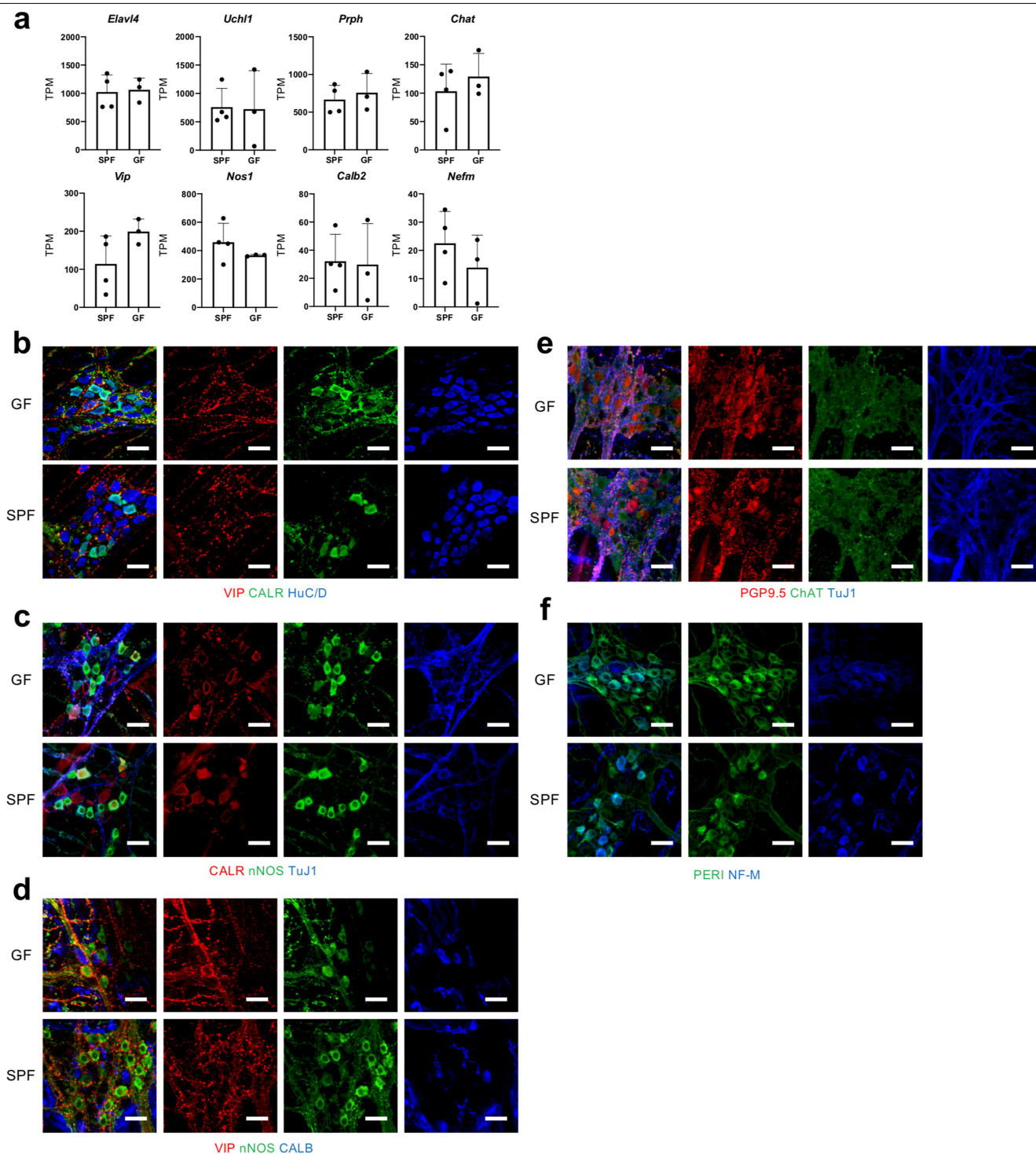
Extended Data Fig. 1 | AAV-based transcriptional profiling of enteric neurons. **a–g**, Representative images of myenteric ganglia from mice injected with the AAV9-CaMKII-eGFP-KASH vector. Colon (**a**) and small intestine (**b–g**) myenteric-plexus preparations were immunostained with antibodies against eGFP (**a–g**), PGP9.5 (**a, b**), S100 β (**c**), SOX10 (**d**), nNOS and calretinin (CALR) (**e**), nNOS and calbindin (CALB) (**f**), and VIP and ChAT (**g**). Data represent two independent experiments. Scale bars, 100 μ m (**a, b**) and 30 μ m (**c–g**). **h**, Percentage of PGP9.5⁺ enteric neurons (mean \pm s.d.) in the proximal small intestine and colon expressing eGFP, following intravenous administration of the AAV9-CaMKII-eGFP-KASH vector. $n = 1,308$ colon neurons and 784 small-intestine neurons from 3 mice. **i**, FACS plots indicating the gating parameters for the isolation of muscularis externa nuclei (gated on DAPI) from the colon (left) and small intestine (right) of mice injected intravenously with AAV9-CaMKII-eGFP-KASH. **j, k**, Peripherin (red) and eGFP (green) whole-mount immunostaining of the colon (**j**) and small intestine (**k**) of mice injected with AAV9-CaMKII-eGFP-KASH mice following dissection of the muscularis externa.

The identification of an intact submucosal plexus demonstrates that our transcriptomic analysis is specific for myenteric neurons. Images represent two independent experiments. Scale bars, 100 μ m. **l**, Volcano plots showing mean log₂-transformed fold change (x axis) and significance ($-\log_{10}$ (adjusted P value)) of differentially expressed genes between eGFP⁺ and eGFP⁻ nuclei isolated from the colon (left) and small intestine (right) of mice injected with AAV9-CaMKII-eGFP-KASH vector. Coloured dots indicate genes specific to enteric neurons (*Ret*, *Chat*, *Camk2a*, *Elavl3*, *Elavl4*, *Nos1* and *Tubb3*) in red, glial cells (*Sox10*, *Gfap*, *Cdh19*, *Entpd2*, *S100b* and *Plp1*)⁴¹ in blue and muscular macrophages (*Itgam*, *Cd163*, *H2-Ab1*, *Mrc1* and *Retnla*)⁴² in green. $n = 4$ mice (Crick). **m**, Principal component analysis of the transcriptomes of eGFP⁺ (neuronal) and eGFP⁻ (non-neuronal) nuclei isolated from the muscularis externa of the colon and small intestine of mice injected with AAV9-CaMKII-eGFP-KASH vectors. Segregation of nuclear transcriptomes according to their neuronal versus non-neuronal origin and anatomical location along the gut. $n = 4$ mice (Crick).



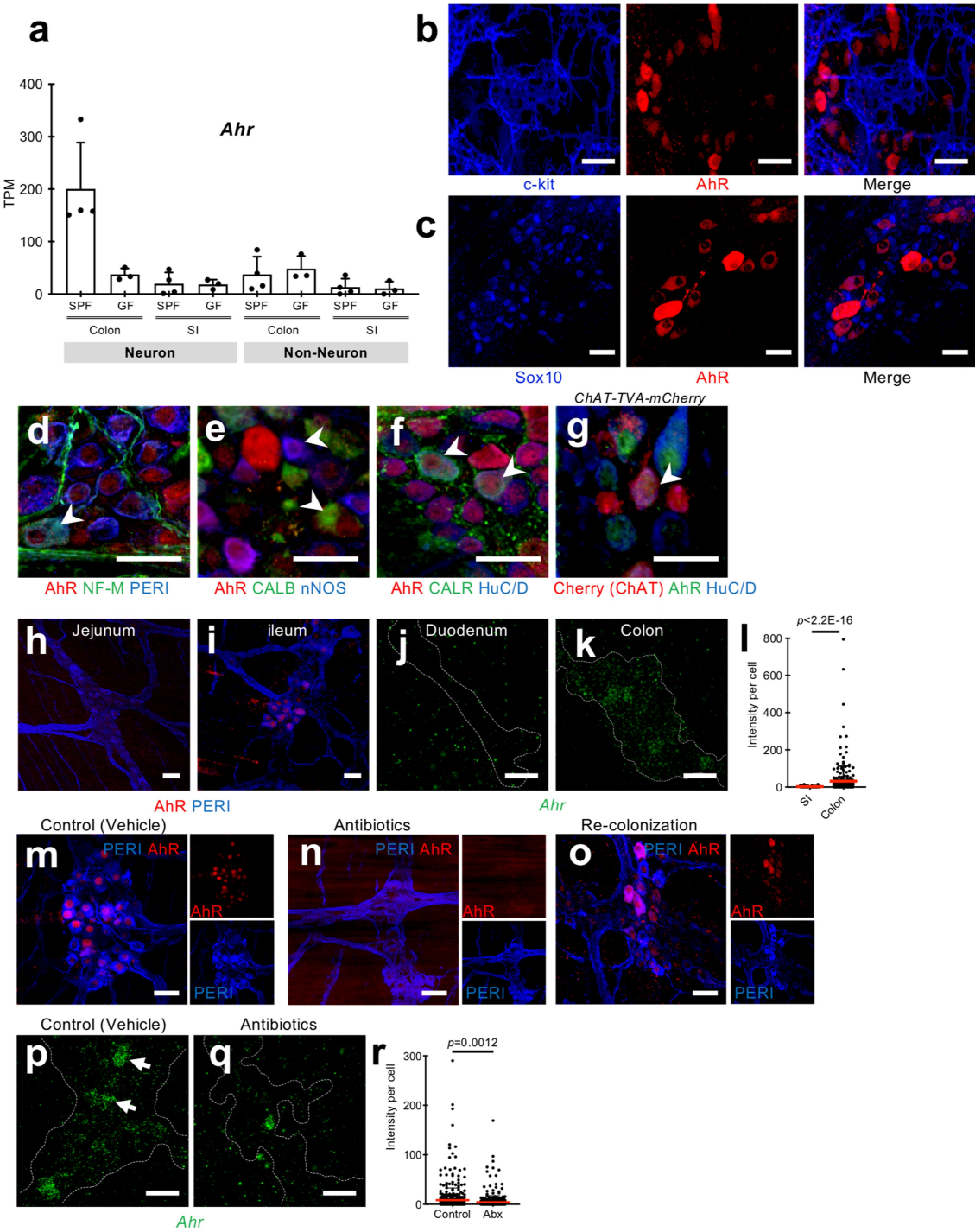
Extended Data Fig. 2 | Differential expression of enteric-neuron-specific genes in myenteric neurons from the small intestine and colon of SPF and germ-free mice. a, b. Representative images of myenteric ganglia (outlined by dotted line) from small intestine (left) and colon (right) of SPF (a) and germ-free (b) mice hybridized with the indicated fluorescence RNAscope probes and counterstained for the pan-neuronal marker HuC/D. *Ret* (positive control for RNAscope detection) is expressed in neurons of myenteric ganglia of both the

small intestine and the colon. *Pou3f3*, *Pde1c*, *Pantr2*, *Ano5*, *Unc5d* and *Col25a1* are expressed at higher levels in colonic versus small intestine neurons in both SPF (a) and germ-free (b) mice. Data represent three independent experiments. Transcripts per kilobase million (TPM) values (mean ± s.d.) for each transcript in small intestine and colon neurons from SPF (a) and germ-free (b) mice. $n = 8$ SPF and 3 germ-free mice. Scale bars, 30 μ m.



Extended Data Fig. 3 | Molecular and neurochemical characterization of colonic neurons in germ-free mice. **a**, TPM values (mean \pm s.d.) for neuronal gene markers *Elavl4*, *Uchl1*, *Prph*, *Chat*, *Vip*, *Nos1*, *Calb2* and *Nefm* in the muscularis externa of the colon of SPF and germ-free mice. $n = 4$ SPF and 3 germ-free mice. **b–f**, Immunostaining of colonic myenteric ganglia from

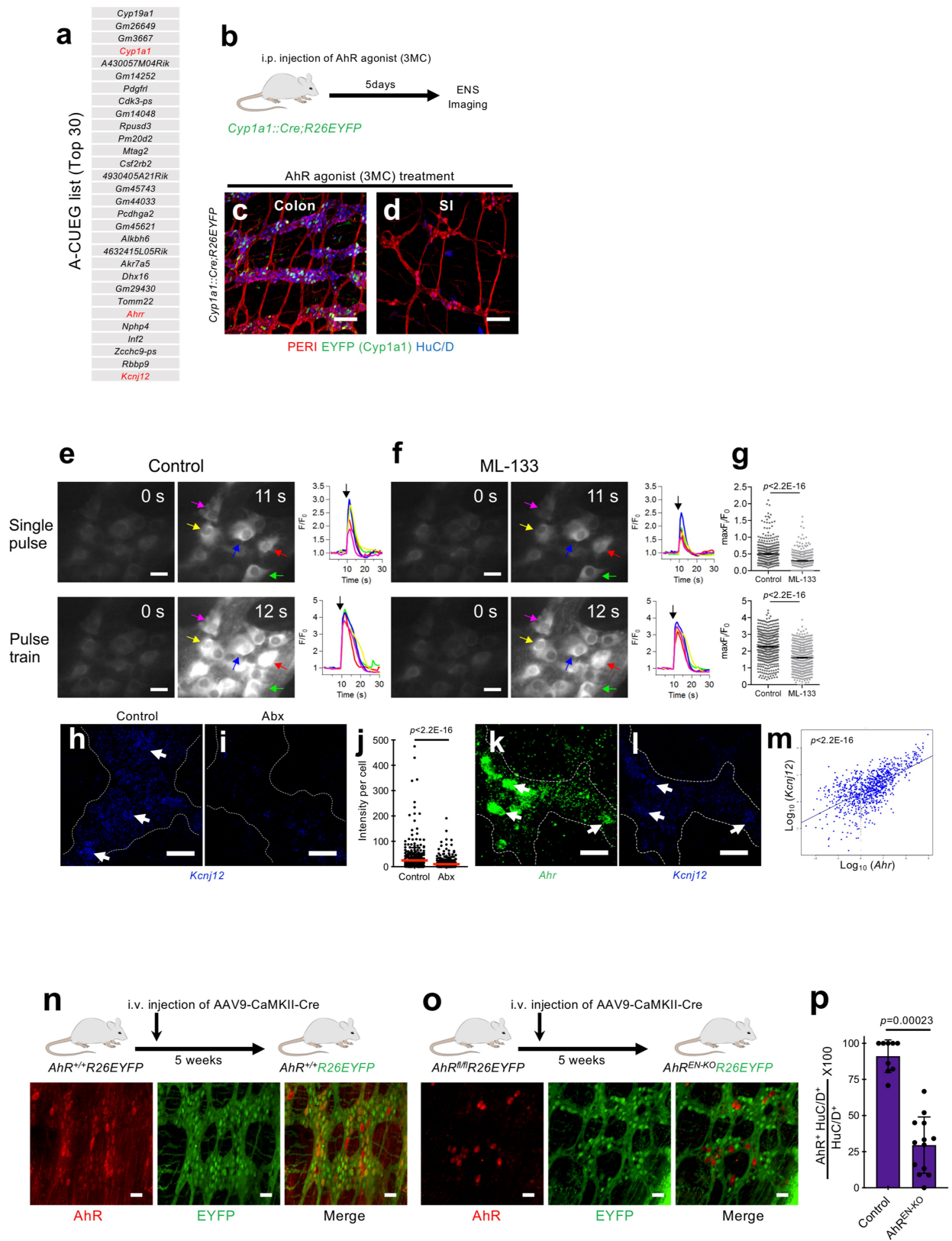
germ-free (top) and SPF (bottom) mice with VIP, CALR and HuC/D (**b**), CALR, nNOS and TuJ1 (**c**), VIP, nNOS and CALB (**d**), PGP9.5, ChAT and TuJ1 (**e**) and peripherin and NF-M (**f**). Scale bars, 30 μ m. Data represent three independent experiments.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Microbiota-dependent expression of AHR in colonic neurons. **a**, TPM values (mean \pm s.d.) for AHR transcripts in neuronal and non-neuronal nuclear preparations from muscularis externa from colon and small intestine of SPF and germ-free mice. $n = 4$ (SPF) and 3 (germ-free) mice. **b, c**, Myenteric ganglia immunostained for KIT (which identifies interstitial cells of Cajal) and AHR (**b**) or SOX10 (enteric glial cells) and AHR (**c**). AHR⁺ cells are distinct from interstitial cells of Cajal and enteric glia. Scale bars, 30 μ m. **d–f**, Immunostaining of neurons from the colon of wild-type mice for AHR (**d–f**) and the neuronal markers peripherin and NF-M (**d**), calbindin and nNOS (**e**), and calretinin and HuC/D (**f**). AHR signal was detected in all subtypes of myenteric neurons (arrowheads). Scale bars, 30 μ m. **g**, Immunostaining of neurons from the colon of *ChAT-mCherry-TVA* reporter mice for mCherry (red), AHR (green) and HuC/D (blue). Arrowhead indicates an enteric neuron positive for ChAT and AHR. Scale bar, 30 μ m. **h, i**, Immunostaining of myenteric ganglia from the jejunum (**h**) and ileum (**i**) with the pan-neuronal marker peripherin (blue) and AHR (red). Scale bar, 30 μ m. **j, k**, Representative images of enteric ganglia from

duodenum (**j**) and colon (**k**) hybridized with RNAscope probe for *Ahr* (green). Dotted line defines the borders of myenteric ganglia. Scale bar, 30 μ m. **l**, Quantification (mean \pm s.d.) of RNAscope signal per neuron is shown (two-sided non-parametric Mann–Whitney *U*-test). $n = 91$ small-intestine and 254 colon neurons from 6 mice. **m–o**, Immunostaining of ganglia from the colon of control (**m**), antibiotic-treated (**n**) and microbiota-colonized, antibiotic-treated (**o**) mice with peripherin (blue) and AHR (red). Small panels show signal for AHR (top) and peripherin (bottom). $n = 3$ mice for each condition. Scale bars, 30 μ m. **p, q**, Representative images of enteric ganglia from the colon of control (**p**) and antibiotic-treated (**q**) mice hybridized with RNAscope probe for *Ahr* (green). Dotted line defines the borders of myenteric ganglia and arrows indicate positive cells. Scale bars, 30 μ m. **r**, Quantification (mean \pm s.d.) of RNAscope signal per neuron is also shown (two-sided non-parametric Mann–Whitney *U*-test). $n = 518$ neurons from 4 control and 468 neurons from 4 antibiotic-treated mice. Data represent two (**j, k, m–q**) or three (**b–i**) independent experiments.



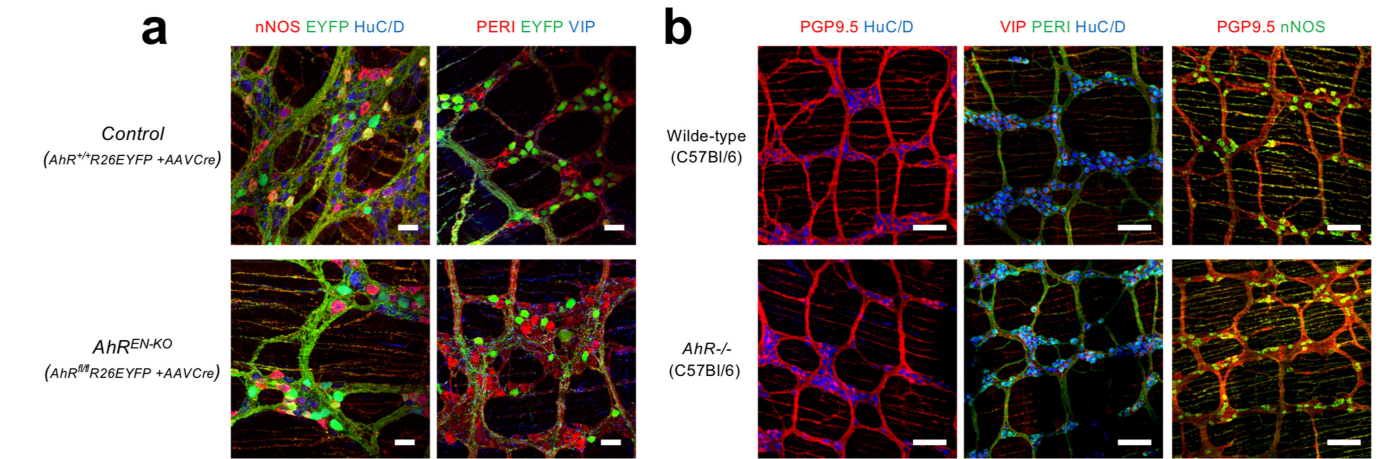
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | AHR-dependent gene expression and effects on colon myenteric neurons. a, The top 30 genes upregulated in colonic neurons by AHR-ligand treatment (AHR-induced CUEGs) were identified on the basis of fold-change criteria (\log_2 -transformed fold change = 2 < maximum).

b, *Cyp1a1::cre;Rosa26eYFP* reporter mice were intraperitoneally injected with 3MC five days before GFP immunostaining. CYP1A1 induction in response to ligand-activated AHR signalling is expected to induce expression of eYFP.

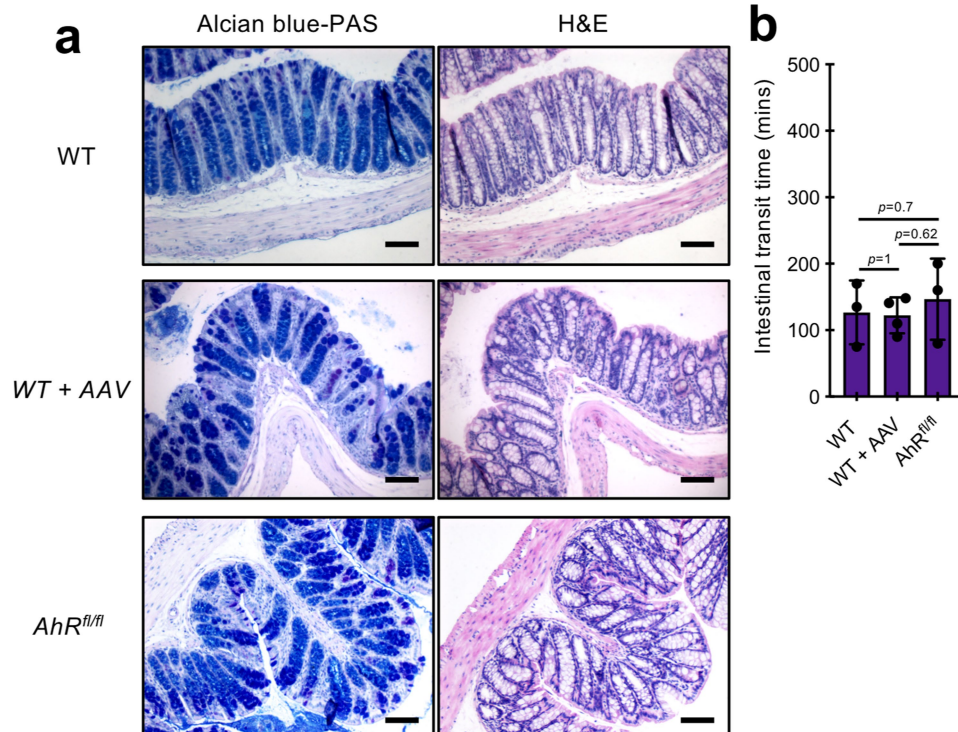
c, d, Immunostaining of myenteric ganglia from the colon (**c**) and small intestine (**d**) of 3MC-treated *Cyp1a1::cre;Rosa26* mice for peripherin (red), HuC/D (blue) and eGFP (green). Scale bars, 100 μ m. Data represent three independent experiments. **e–g,** Live calcium imaging of colonic myenteric plexus preparations from *Wnt1::cre;Rosa26-GCaMP6f* mice. Electrically stimulated Ca^{2+} transients in enteric neurons under control conditions (**e**) or in the presence of the ML-133 blocker²¹ (10 μ M) (**f**). Data represent four independent experiments. The greyscale images depict a proximal colon myenteric plexus preparation in which enteric neurons were stimulated by a single electrical pulse (top panels) or an electrical pulse train (1 s, 20 Hz; bottom panels) via a focal electrode positioned on an internodal strand leading into the myenteric ganglion in the field of view. Left, baseline before stimulation. Middle, peak GCaMP6f fluorescence of the same ganglion upon electrical stimulation. Scale bars, 20 μ m. Right, Ca^{2+} transients of individual enteric neurons (indicated by colour-coded arrows shown in the middle panels) induced by electrical stimulation. The electrical stimulus was applied at 10 s as marked by the black arrows. Comparison of the average maximal GCaMP6f fluorescence amplitudes of neuronal Ca^{2+} responses (mean \pm s.e.m.) under control conditions (**e**) or the presence of ML-133 (**f**) upon single pulse (top

($n = 457$ neurons) and pulse train (bottom) ($n = 526$ neurons) electrical stimulation is shown in **g** (two-sided paired t -test). **h, i,** Myenteric ganglia from colon of control (**h**) and antibiotic-treated (**i**) mice hybridized with the *Kcnj12* RNAscope probe. Dotted line defines the borders of myenteric ganglia and arrows indicate *Kcnj12*-expressing cells. Scale bars, 30 μ m. Data represent two independent experiments. **j,** Quantification of RNAscope signal (mean \pm s.e.m.) shown in **h** and **i** (two-sided non-parametric Mann–Whitney U -test). $n = 421$ neurons from 4 control and 468 neurons from 4 antibiotic-treated mice. Abx, antibiotics. **k, l,** Myenteric ganglia from colon of SPF mice hybridized with the *Ahr* (green) (**k**) and *Kcnj12* (blue) (**l**) RNAscope probes and immunostained with HuC/D (data not shown). Dotted line defines the borders of myenteric ganglia and arrows indicate AHR- and KCNJ12-expressing neurons. Scale bars, 30 μ m. Data represent two independent experiments. **m,** Scatter plot shows positive correlation in RNAscope signal for *Ahr* (**k**) and *Kcnj12* (**l**) in myenteric neurons (F -test). $n = 1,037$ neurons from 3 mice. **n, o,** Immunostaining of myenteric ganglia from control (*Ahr*^{+/+}; *Rosa26eYFP* injected with the AAV9-CaMKII-Cre vector) (**n**) and *Ahr*^{EN-KO} (**o**) mice for AHR (red) and eYFP (green). Note the lack of overlap between green and red signal in the case of *Ahr*^{EN-KO} (**o**). Data are representative of two independent experiments. Scale bars, 30 μ m. **p,** Percentage of AHR⁺ neurons in myenteric ganglia of control (*Ahr*^{+/+}; *Rosa26eYFP* mice injected with the AAV9-CaMKII-Cre vector) and *Ahr*^{EN-KO} mice. Random images were acquired from the colon of each biological replicate ($n = 9$ for control, $n = 13$ for *Ahr*^{EN-KO}), and the average percentage (mean \pm s.d.) of AHR⁺ HuC/D⁺ cells among the total population of HuC/D⁺ neurons was calculated (two-sided Student's t -test).



Extended Data Fig. 6 | Deletion of *Ahr* does not alter the organization and composition of myenteric ganglia. a, Immunostaining of muscularis externa preparations from the colon of control (top) and *Ahr*^{REN-KO} (bottom) mice with nNOS, eYFP and HuC/D (left) or peripherin, eYFP and VIP (right). Scale bars,

30 μm. **b,** Immunostaining of muscularis externa preparations from the colon of wild-type (top) and *Ahr*^{-/-} (bottom) mice with PGP9.5 and HuC/D (left), VIP, peripherin and HuC/D (middle) and PGP9.5 and nNOS (right). Scale bars, 100 μm. Data represent three independent experiments.



Extended Data Fig. 7 | Intravenous administration of AAV vectors does not elicit an inflammatory response or intestinal dysmotility. **a**, Cross-sections from the colon of wild-type (top), wild-type infected with AAV9-CaMKII-Cre (middle) and *Ahr^{fl/fl}* (bottom) mice stained with Alcian blue-PAS (left) or haematoxylin and eosin (H&E) (right). Data represent two independent

experiments. **b**, Graph (mean \pm s.d.) shows that administration of AAV-CaMKII-Cre vector into wild-type mice is not sufficient to alter intestinal transit time. $n = 3$ (wild type), 4 (WT + AAV) or 3 (*Ahr^{fl/fl}*) mice. Statistical test is a two-sided non-parametric Mann-Whitney *U*-test. Scale bars, 50 μ m.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

TillVision (Till Photonics, Gräfelfing, DE), Olympus FV3000, BD FACSDiva, Applied Biosystems™ QuantStudio™ 12K Flex Real-Time PCR System, Applied Biosystems™ QuantStudio™ 7 Flex Real-Time PCR System

Data analysis

Igor Pro. 8.02 (Wavemetrics, Lake Oswego, Oregon, US), R v3.3.1, Prism 8, FlowJo (version 10.5.3), Fiji/ImageJ (Wayne Rasband, NIH), Testing and Timing computer is a Dell Precision 7820 Workstation, Intel Xeon Silver 4114 @2.20 GHz (40 CPUs), 132 Gb memory. The OS is Windows 10, Cell Profiler Version 3.1.8, MATLAB 2018b. Trim Galore! Version 0.4.2, RSEM Version 1.3.0, STAR Version 2.5.2, MultiQC version 1.3, R Version 3.3.1, R-packages: DESeq2, ggplot2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All RNA-seq data are available at Gene Expression Omnibus under accession number GSE140293

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined based on preliminary experiments or reports in the literature
Data exclusions	The intestinal transit time (ITT) data from one animal was excluded due to health issue.
Replication	Experiments were repeated with at least two biologically independent experiments. In all cases results were reproducible.
Randomization	Littermates (of mixed or the same sex) were randomly assigned to experimental groups in an age range of 8-16 weeks unless otherwise specified.
Blinding	Blinding was performed wherever possible, such as experiments in Fig 3a-d and h-k, Extended Data Fig. 7c and Extended Data Fig. 9b

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

α-AhR Enzo Life Sciences BML-SA550-0100 Rabbit PC 1:200
 α-cKit R&D Systems AF1356 Goat PC 1:200
 α-Calbindin Abcam ab82812 Mouse MC 1:200
 α-Calretinin Swant CG1 Goat PC 1:200
 α-ChAT Millipore (Chemicon) AB144P Goat PC 1:200
 α-GFP Abcam ab13970 Chicken PC 1:200
 α-HuC/D Invitrogen A-21271 Mouse MC 1:200
 α-mCherry SICGEN AB0040-200 Goat PC 1:200
 α-NF-M Proteintech 66396-1-Ig Mouse MC 1:200
 α-nNOS Abcam ab1376 Goat PC 1:200
 α-Peripherin Santa Cruz sc-7604 Goat PC 1:200
 α-PGP9.5 Bio-Rad 7863-0504 Rabbit PC 1:1000
 α-S100β DAKO z-0311 Rabbit PC 1:500
 α-Sox10 Santa-Cruz Biotechnology sc-17342 Goat PC 1:200
 α-Tuj1 Biolegend MMS-435P (801201) Mouse MC 1:200
 α-VIP Abcam ab22736 Rabbit PC 1:200

 α-Mouse Alexa Fluor 405 Abcam ab175658 Donkey PC 1:500
 α-Mouse Alexa Fluor 488 Invitrogen (Life Technologies) A-21202 Donkey PC 1:500
 α-Goat Alexa Fluor 488 Invitrogen (Life Technologies) A-11055 Donkey PC 1:500
 α-Rabbit Alexa Fluor 568 Invitrogen (Life Technologies) A-10042 Donkey PC 1:500
 α-Goat Alexa Fluor 568 Invitrogen (Life Technologies) A-11057 Donkey PC 1:500
 α-Mouse Alexa Fluor 647 Invitrogen (Life Technologies) A-31571 Donkey PC 1:500
 α-Rabbit AlexaFluor 647 Invitrogen (Life Technologies) A-31573 Donkey PC 1:500
 α-Goat Alexa Fluor 647 Invitrogen (Life Technologies) A-21447 Donkey PC 1:500

α-Rabbit Alexa Fluor 488 Invitrogen (Life Technologies) R37118 Donkey PC 1:500
 α-Chicken Alexa Fluor 488 Jackson ImmunoResearch 703-545-155 Donkey PC 1:500

Validation

All antibodies used in our study are commercially available and have been validated by the manufacturer. The catalog number of each antibody is described above.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

C57BL/6, Cyp1a1::Cre, AhR^{-/-}, AhR^{fl/fl}, R26LSL-Cyp1a, R26EYFP, Wnt1::Cre, Ai95D(RCL-GCaMP6f)-D. and ChAT-TVA-mCherry mice were used in this study. For our experiments we used male mice unless stated otherwise in Methods or relevant figure legends. Animals were age matched and used at 8-16 weeks of age. All mice were maintained under specific pathogen-free (SPF) conditions on a 12-hour light/dark cycle (7am-7pm), and provided food and water ad libitum. For the intestinal motility rescue experiment, animals were fed with purified diet supplemented with I3C (200 mg/kg; Sniff Spezialdiäten, GmbH). GF mice were maintained in flexible-film isolators at the Clean Mouse Facility of the University of Bern (Switzerland). For bacterial colonisation experiments, faecal contents of SPF mice were orally administered to wild-type GF mice and colonised animals were co-housed and maintained in the SPF facility of the University of Bern for 4 weeks prior the analysis.

Wild animals

No wild animals included.

Field-collected samples

No field-collected samples included.

Ethics oversight

All work was carried under a Home Office Project Licence and was approved by the Crick Institute's Animal Welfare and Ethical Review Body (UK). Animal work at the University of Bern was carried out according to Swiss Federal Regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

described in detail in materials and methods section.

Instrument

FACS Aria Fusion

Software

DIVA and FLOWJO software (latest version)

Cell population abundance

determined by expression analysis (QPCR) of sorted nuclei

Gating strategy

we used negative control (AAV-non-injected) sample for negative gating

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

The guidance receptor plexin D1 is a mechanosensor in endothelial cells

<https://doi.org/10.1038/s41586-020-1979-4>

Received: 9 November 2018

Accepted: 5 December 2019

Published online: 5 February 2020

Vedanta Mehta^{1,2}, Kar-Lai Pang^{1,2}, Daniel Rozbesky^{2,3}, Katrin Nather^{1,2}, Adam Keen^{1,2}, Dariusz Lachowski⁴, Youxin Kong^{2,3}, Dimple Karia^{2,3}, Michael Ameismeier^{2,3}, Jianhua Huang⁵, Yun Fang⁶, Armando del Rio Hernandez⁴, John S. Reader^{1,2,7}, E. Yvonne Jones^{2,3,7} & Ellie Tzima^{1,2,7*}

Shear stress on arteries produced by blood flow is important for vascular development and homeostasis but can also initiate atherosclerosis¹. Endothelial cells that line the vasculature use molecular mechanosensors to directly detect shear stress profiles that will ultimately lead to atheroprotective or atherogenic responses². Plexins are key cell-surface receptors of the semaphorin family of cell-guidance signalling proteins and can regulate cellular patterning by modulating the cytoskeleton and focal adhesion structures^{3–5}. However, a role for plexin proteins in mechanotransduction has not been examined. Here we show that plexin D1 (PLXND1) has a role in mechanosensation and mechanically induced disease pathogenesis. PLXND1 is required for the response of endothelial cells to shear stress in vitro and in vivo and regulates the site-specific distribution of atherosclerotic lesions. In endothelial cells, PLXND1 is a direct force sensor and forms a mechanocomplex with neuropilin-1 and VEGFR2 that is necessary and sufficient for conferring mechanosensitivity upstream of the junctional complex and integrins. PLXND1 achieves its binary functions as either a ligand or a force receptor by adopting two distinct molecular conformations. Our results establish a previously undescribed mechanosensor in endothelial cells that regulates cardiovascular pathophysiology, and provide a mechanism by which a single receptor can exhibit a binary biochemical nature.

Endothelial cells (ECs) are constantly exposed to the haemodynamic forces of blood flow, including the frictional force of fluid shear stress that—depending on vessel geometry—can be protective or pathogenic. Whereas disturbed or atheroprone flow patterns found in curvatures and bifurcations are associated with the upregulation of pro-inflammatory genes and the deposition of atherosclerotic lesions, uniform or atheroprotective shear stress induces the remodelling of the cytoskeleton and the alignment of ECs in the direction of flow^{1,6}. The importance of shear stress in the development and function of the cardiovascular system has inspired efforts to identify endothelial mechanosensors, as they are the first to respond to changes in the mechanical environment².

Plexins are cellular receptors that have a range of important functions in axon guidance, tumour progression and immune-cell regulation⁷. Plexins are known to act primarily by binding to semaphorin ligands, in a cell-bound or free state in solution along with other coreceptors, resulting in intracellular signalling events that lead to large-scale changes in the cytoskeleton and cell adhesion^{3,4}. Here we show that the guidance receptor PLXND1 acts as a mechanosensor in ECs, regulating vascular function and the site-specific distribution of atherosclerosis.

To determine the role of PLXND1 under flow conditions, we transfected bovine aortic ECs with either a scrambled short interfering RNA (siRNA) or siRNA against *PLXND1* (Extended Data Fig. 1a) and subjected the cells to shear stress. Knockdown of *PLXND1* attenuated the activation induced by shear stress of the key signalling mediators Akt, ERK1 and ERK2 (hereafter ERK1/2) and eNOS (Extended Data Fig. 2a). PLXND1-dependent mechanotransduction is independent of its ligand SEMA3E, as incubation with a SEMA3E-blocking antibody did not affect the flow-induced activation of signalling cascades (Extended Data Fig. 3). Next, we examined the role of PLXND1 in the hallmark response to atheroprotective shear stress by examining the alignment of ECs in the direction of flow. EC alignment with flow direction is highly correlated with atheroresistant regions of arteries and has an important function in the activation of anti-inflammatory pathways. PLXND1-depleted bovine ECs showed a notable failure to align in response to shear stress, and displayed fewer and more-disorganized actin stress fibres (Extended Data Fig. 2b). Quantification of the EC alignment by measuring the orientation angle and the elongation factor indicate that PLXND1 is required for the alignment of ECs with the flow. We also examined the mRNA levels of the Kruppel-like factors *Klf2* and *Klf4*—key anti-inflammatory

¹Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ²Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. ³Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. ⁴Cellular and Molecular Biomechanics Laboratory, Department of Bioengineering, Imperial College London, London, UK. ⁵Department of Medicine, Duke University, Durham, NC, USA. ⁶Department of Medicine, University of Chicago, Chicago, IL, USA. ⁷These authors jointly supervised this work: John S. Reader, E. Yvonne Jones, Ellie Tzima. *e-mail: ellie@well.ox.ac.uk

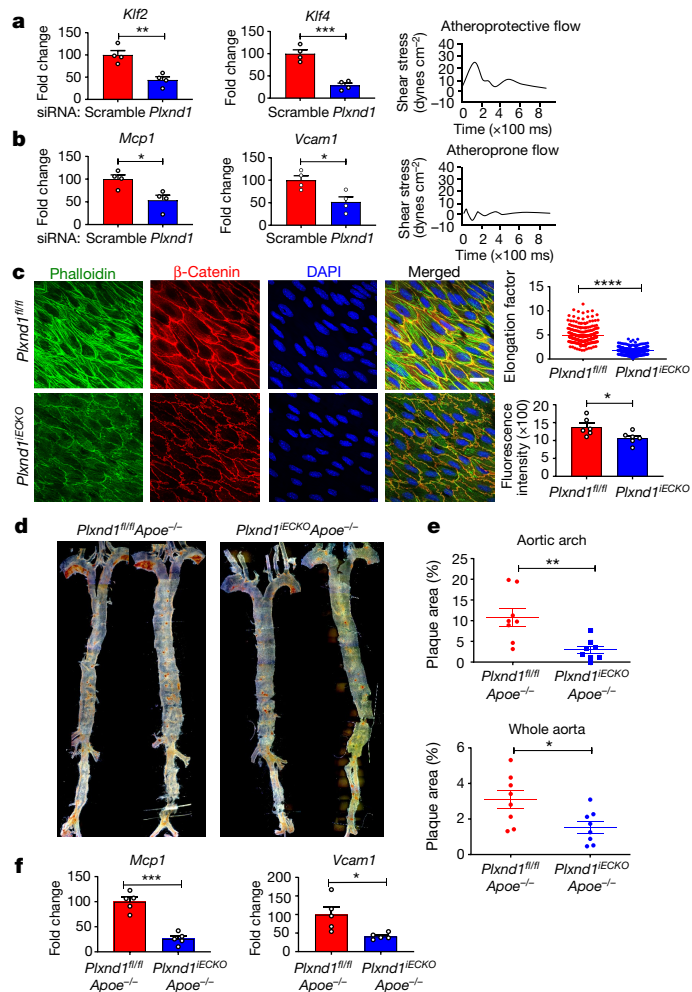


Fig. 1 | PLXND1 mediates the EC response to fluid shear stress and regulates the site-specific distribution of atherosclerosis. **a, b,** Mouse ECs were

transfected with either scrambled or *Plxnd1* siRNA and exposed to either atheroprotective or atheroprone flow for 24 h, using a cone-and-plate viscometer. qPCR was performed to quantify the expression of *Klf2* and *Klf4* in samples subjected to atheroprotective flow, and expression of the inflammatory markers *Mcp1* (also known as *Ccl2*) and *Vcam1* in samples subjected to atheroprone flow. *n* = 4 biological replicates. **c,** The descending thoracic aorta was isolated and prepared en face from *Plxnd1*^{fl/fl} and *Plxnd1*^{IECKO} mice and stained for β -catenin, phalloidin and DAPI to visualize the cell junctions, actin stress fibres and nuclei. Quantification of alignment and mean fluorescence intensity was performed using ImageJ; 3–5 images (each image has *n* ≤ 100 cells) taken from 3 regions along the length of the descending aorta were obtained from *n* = 5 mice of each genotype (exact sample numbers are provided in the Source Data). **d,** Representative en face preparations of whole aortas showing atherosclerosis in *Plxnd1*^{fl/fl} *Apoe*^{-/-} and *Plxnd1*^{IECKO} *Apoe*^{-/-} mice after 10 weeks of high-fat diet feeding, visualized by oil-Red-O staining. **e,** Quantification of the lesion area in whole aortas and aortic arches from *Plxnd1*^{fl/fl} *Apoe*^{-/-} and *Plxnd1*^{IECKO} *Apoe*^{-/-} mice. *n* = 8 mice. **f,** Aortic arches from *Plxnd1*^{fl/fl} *Apoe*^{-/-} and *Plxnd1*^{IECKO} *Apoe*^{-/-} mice were isolated and qPCR was performed for expression of the inflammatory markers *Vcam1* and *Mcp1*. *n* = 5. Data are mean ± s.e.m. *P* values were obtained using two-tailed Student's *t*-tests using GraphPad Prism. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. Scale bar, 20 μ m.

transcription factors that are known to be upregulated by atheroprotective shear stress^{8,9}. We found that knockdown of PLXND1 attenuated flow-induced upregulation of both of these genes compared with control mouse ECs (Fig. 1a). We then investigated whether PLXND1 could mediate the endothelial response to disturbed shear stress.

We subjected mouse ECs to atheroprone flow for 24 h and examined mRNA levels of the pro-inflammatory genes monocyte chemoattractant protein-1 (*Mcp1* (also known as *Ccl2*)) and vascular cell adhesion molecule-1 (*Vcam1*)¹⁰. We noted that knockdown of PLXND1 in ECs treated with siRNA significantly reduced the upregulation of both genes in response to atheroprone shear stress (Fig. 1b). Taken together, these data demonstrate that PLXND1 is a critical mediator of key shear-stress responses in ECs.

To explore the biological relevance of our findings, we used a transgenic mouse model to enable the endothelial-specific inducible deletion of PLXND1 (*Plxnd1*^{IECKO}) (Extended Data Figs. 1b, 7b). Confocal imaging of actin filaments in ECs and staining for the junctional marker β -catenin revealed a reduction in the elongation of ECs and a reduced intensity of actin stress fibres in the absence of PLXND1 (Fig. 1c)—consistent with in vitro observations (Extended Data Fig. 2b).

Given the decrease in the expression of inflammatory genes in response to atheroprone shear stress in vitro that was observed with loss of PLXND1 (Fig. 1b), we assessed the role of endothelial PLXND1 in a pathophysiological setting. Atherosclerotic lesions are known to occur in regions of the vasculature with low or disturbed blood flow, flow reversal and other complex spatiotemporal flow patterns¹. Systemic risk factors, such as hypercholesterolaemia, interact with local biomechanical factors to initiate and advance the deposition of atherosclerotic plaques. To assess whether endothelial deletion of PLXND1 affected atherosclerosis in vivo, we crossed *Plxnd1*^{fl/fl} and *Plxnd1*^{IECKO} mice with hypercholesterolaemic apolipoprotein-E deficient (*Apoe*^{-/-}) mice¹¹ and fed them a high-fat diet for 10 weeks. Although the body weights and lipid levels of the mice were unaffected by the loss of PLXND1 (Extended Data Fig. 4a), quantification of oil-red-O-stained aortic samples revealed a significant decrease in the plaque burden of both the whole aorta and the aortic arch in *Plxnd1*^{IECKO} *Apoe*^{-/-} mice (Fig. 1d, e). To explore these differences further, we examined the expression of inflammatory markers in the inner curvature of the aortic arch. Immunostaining and quantitative (q)PCR analysis showed reduced levels of MCP-1 and VCAM-1 in *Plxnd1*^{IECKO} *Apoe*^{-/-} compared with *Plxnd1*^{fl/fl} *Apoe*^{-/-} mice (Fig. 1f and Extended Data Fig. 4b). Given the atheroprotective role of laminar shear stress and the reduced alignment of ECs with loss of PLXND1, we examined the effects in the atheroprotected descending aorta. After a high-fat diet for an extended period of 20 weeks, we observed an increase in the plaque burden in the descending aortas of *Plxnd1*^{IECKO} *Apoe*^{-/-} mice (Extended Data Fig. 5); these plaques also appeared to correlate with intercostal branch points that have flow disturbances. Together, these results show that endothelial PLXND1 is required for the endothelial response to fluid shear stress and the site-specific distribution of atherosclerosis.

The requirement of PLXND1 in flow-mediated responses in vitro and in vivo prompted us to investigate whether this is because PLXND1 is simply a player in mechanochemical signalling cascades or functions as a mechanoreceptor that is capable of detecting mechanical force. We applied tensional forces, using a magnetic system¹², to paramagnetic beads coated with an antibody that recognizes the extracellular domain of PLXND1 and examined force responses using four different readouts. First, force on PLXND1 induced activation of the same signalling cascades (ERK1/2, Akt and VEGFR2) (Fig. 2a), as those induced by shear stress¹³ (Extended Data Fig. 2a). Second, we observed a robust transient increase in intracellular calcium levels in ECs when force was applied on PLXND1 (Fig. 2b), similar to the response observed for other recently discovered mechanosensors^{14,15}. Third, we examined cytoskeletal responses¹²: ECs responded to the application of force on PLXND1 by exhibiting a robust increase in both vinculin-positive focal adhesions (Fig. 2c) and ligated integrin β 1 staining (Extended Data Fig. 6a). Notably, the mechanotransduction response was not restricted to the vicinity of the magnetic bead under tension, but

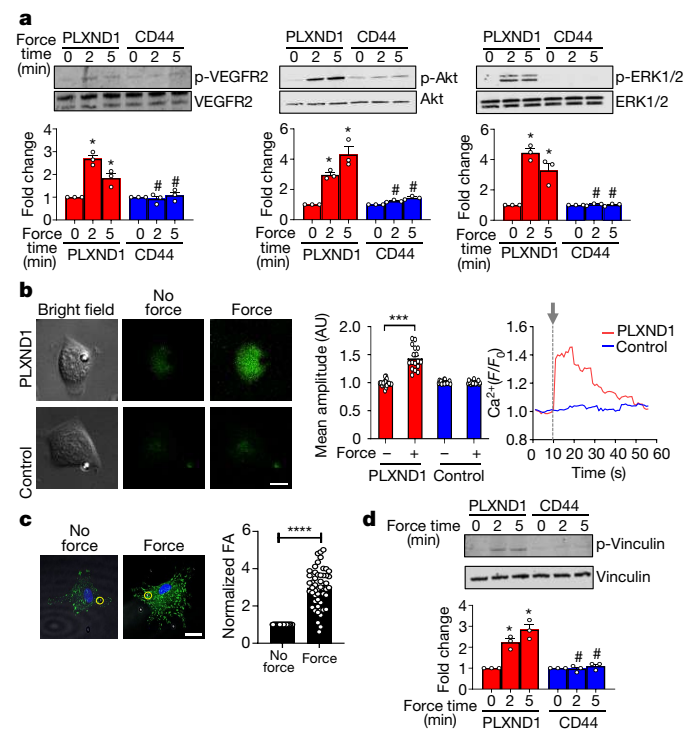


Fig. 2 | PLXND1 is a mechanosensor that mediates the EC response to force.

a, Mouse ECs were incubated with anti-PLXND1 or C44 (negative control) antibody-coated beads and subjected to force (10 pN) for the indicated time periods. Phosphorylation of VEGFR2, Akt and ERK1/2 was determined by western blotting and quantified using Image Studio Lite v.5.2. $n = 3$ biological repeats. Phosphorylated proteins are indicated by 'p'. * $P < 0.05$ relative to the no force condition, # $P < 0.05$ relative to the respective force application time point with PLXND1. **b**, Bovine aortic ECs were loaded with Fluo-8AM dye and then incubated with beads coated with an antibody against the extracellular domain of PLXND1 or poly-L-lysine (negative control). The beads were then subjected to force (1 nN). Calcium responses were measured by calculating the fluorescent intensity of individual cells before (10 s), during (20 s) and after (30 s) stimulation. Representative images are shown along with quantification. $n = 18$ cells for PLXND1 and $n = 19$ cells for control across 3 independent biological replicates. *** $P < 0.001$ relative to unstimulated controls. Scale bar, 10 μm . A representative trace of the calcium influx response over time is also shown. The arrow marks the start of the stimulation. AU, arbitrary units. **c**, Bovine aortic ECs were incubated with anti-PLXND1-coated beads and subjected to force (10 pN) for 30 min. ECs were fixed and stained with an anti-vinculin antibody to mark focal adhesions. Focal adhesion (FA) numbers were quantified using ImageJ software. Values were normalized to the no force condition. Locations of the beads are highlighted in yellow circles. $n = 50$ cells for each condition from 3 independent biological replicates. **** $P < 0.0001$. Scale bar, 10 μm . **d**, Mouse ECs were incubated with anti-PLXND1 or CD44 antibody-coated beads and subjected to 10 pN force for the indicated time periods. Phosphorylation of vinculin was determined by western blotting and quantified using Image Studio Lite v.5.2. $n = 3$ biological repeats. * $P < 0.05$ relative to the no force condition, # $P < 0.05$ relative to the respective force application time point with PLXND1. Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -tests using GraphPad Prism.

was a global cell-wide phenomenon. Fourth, we examined the phosphorylation of vinculin at Y822, a site known to be phosphorylated when force is applied on E-cadherin¹⁶, and observed a significant increase in its activation after force application on PLXND1 (Fig. 2d). These mechanoresponses were specific to PLXND1, as ECs incubated with beads coated with another transmembrane receptor (CD44) or poly-L-lysine did not respond to force. These data demonstrate that PLXND1 is a direct force sensor that can elicit robust and global mechanical signalling in ECs.

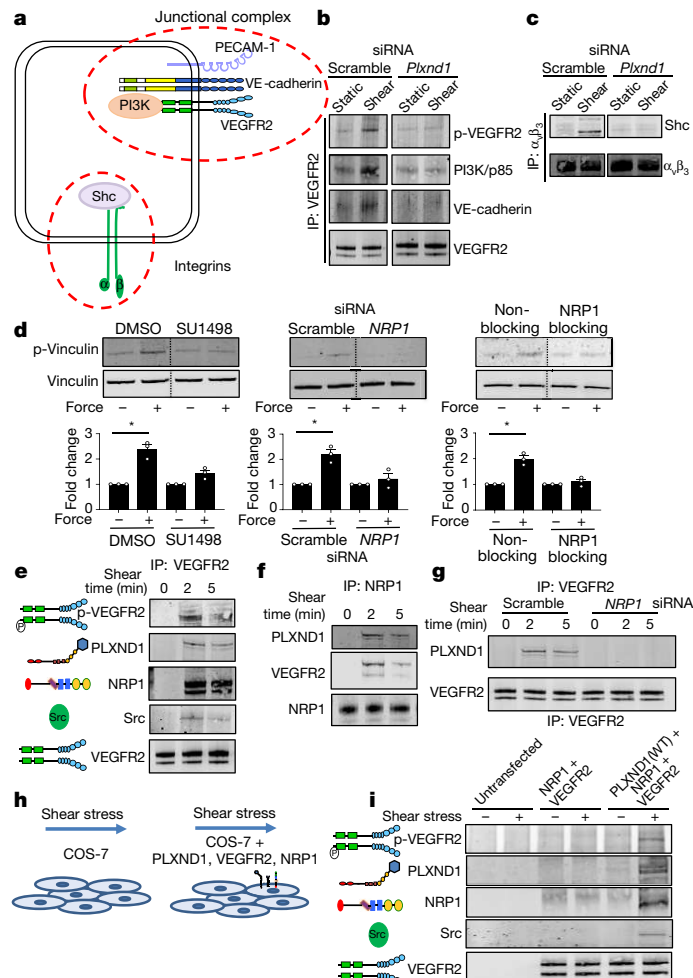


Fig. 3 | The PLXND1, NRP1 and VEGFR2 mechanocomplex functions upstream of known mechanosensory hotspots and is sufficient for the response to shear stress. **a**, Schematic showing signalling at the junctional complex and integrins. **b**, Mouse ECs were transfected with scrambled or *Plxnd1* siRNA, exposed to shear stress for 2 min before immunoprecipitating VEGFR2 and analysis of the phosphorylation and association of VEGFR2 with the p85 subunit of PI3K and VE-cadherin. $n = 3$. IP, immunoprecipitation. **c**, Mouse ECs were transfected with scrambled or *Plxnd1* siRNA, exposed to shear stress for 30 min before immunoprecipitating integrin $\alpha_v\beta_3$ and analysis of the association of integrin $\alpha_v\beta_3$ with Shc. $n = 3$. **d**, Mouse ECs were treated with the VEGFR2 kinase inhibitor SU1498, transfected with siRNA against *Nrp1* or treated with a NRP1-blocking antibody, incubated with anti-PLXND1-coated beads and subjected to force for 5 min before analysis of the phosphorylation of vinculin. ($n = 3$; * $P < 0.05$) Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -test using GraphPad Prism. **e**, Mouse ECs were exposed to shear stress before immunoprecipitating VEGFR2 and analysis of the phosphorylation and association of VEGFR2 with PLXND1, NRP1 and Src. $n = 3$. **f**, Mouse ECs were exposed to shear stress before immunoprecipitating NRP1 and analysis of the association of NRP1 with PLXND1 and VEGFR2. $n = 3$. **g**, Mouse ECs transfected with either scrambled or *Nrp1* siRNA were exposed to shear stress before immunoprecipitating VEGFR2 and analysis of the association of VEGFR2 with PLXND1. $n = 3$. **h**, Schematic showing that reconstitution of PLXND1, VEGFR2 and NRP1 in COS-7 cells confers shear-stress sensitivity to these cells. **i**, COS-7 cells were left untransfected or transfected with NRP1 and VEGFR2, with or without PLXND1 before being subjected to shear stress for 2 min and VEGFR2 was immunoprecipitated. Shear-stress sensitivity was assessed by analysing the levels of phosphorylated VEGFR2, the complex formation between VEGFR2 and Src and the complex formation of PLXND1, VEGFR2 and NRP1. $n = 3$. All shear-stress experiments were at 12 dynes cm^{-2} using a parallel plate system.

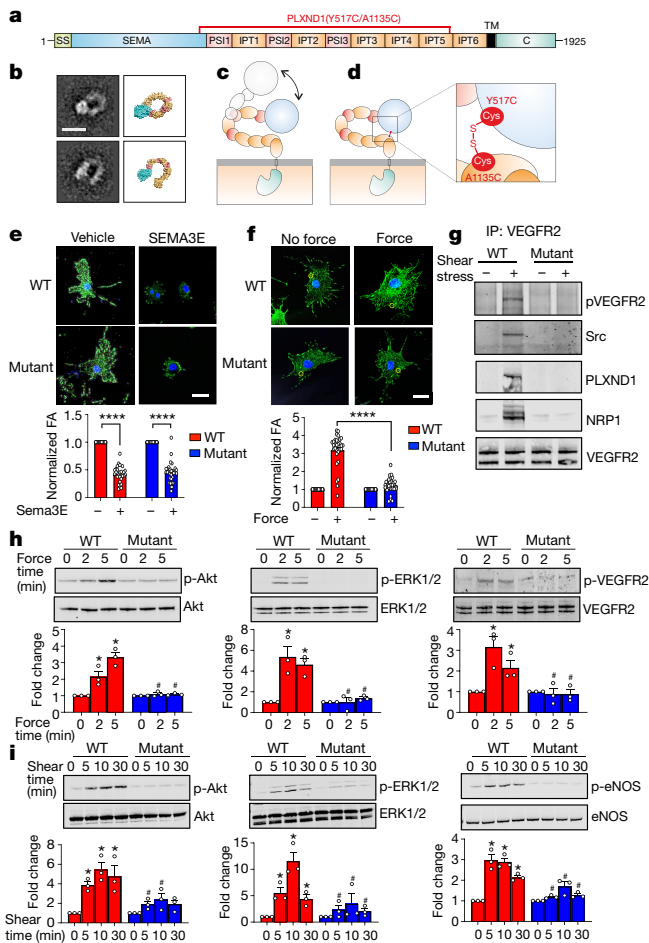


Fig. 4 | PLXND1 flexion is required for mechanotransduction. **a**, Schematic domain organization of PLXND1 spanning amino acids 1–1925. SS, signal sequence; TM, transmembrane region; c, cytoplasmic region. **b**, Representative negative-stain class averages of the PLXND1 ectodomain and corresponding structural models showing the ring-like and open conformations. Scale bar, 10 nm. Two-dimensional class averages were obtained by classifying 1,357 particles into 10 classes. **c**, Model of opening of the ring-like ectodomain, which confers the mechanosensory functions of the PLXND1. **d**, Design of the PLXND1 mutant with an intramolecular disulfide bond to lock the ring-like structure. The magnified view shows the disulfide bond between the SEMA domain (domain 1) and IPT5 domain (domain 9). **e**, **f**, Bovine ECs in which endogenous PLXND1 was knocked down were infected with adenoviruses expressing wild-type or mutant PLXND1, treated with SEMA3E for 30 min or incubated with anti-PLXND1 paramagnetic beads followed by force application (10 pN; 30 min). Cells were immunostained with anti-vinculin antibodies. Focal adhesion numbers were quantified using ImageJ; $n = 30$ cells across either 4 (**e**) or 3 (**f**) biological replicates. **** $P < 0.0001$. Scale bar, 10 μ m. **g**, COS-7 cells were transfected with wild-type or mutant PLXND1, NRP1 and VEGFR2 before shear-stress application for 2 min and VEGFR2 was immunoprecipitated. Shear-stress sensitivity was assessed by analysing the levels of phosphorylated VEGFR2, the complex formation between VEGFR2 and Src and the complex of PLXND1, VEGFR2 and NRP1. $n = 3$. **h**, Mouse ECs in which endogenous PLXND1 was knocked down were infected with adenoviruses expressing wild-type or mutant PLXND1 and incubated with anti-PLXND1 paramagnetic beads followed by force application (10 pN). Phosphorylation of Akt, ERK1/2 and VEGFR2 was determined. $n = 3$. * $P < 0.05$ relative to the no force condition; # $P < 0.05$ relative to the force time point of the respective wild-type protein. **i**, Mouse ECs in which endogenous PLXND1 was knocked down were infected with adenoviruses expressing wild-type or mutant PLXND1 and subjected to fluid shear stress. Phosphorylation of Akt, ERK1/2 and eNOS was determined. $n = 3$ biological repeats. * $P < 0.05$ relative to the static condition; # $P < 0.05$ relative to the respective shear time point of the wild-type protein. Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -test using GraphPad Prism.

Cell–cell and cell–matrix adhesions represent two sites that are highly mechanically active within ECs. These sites include the junctional mechanosensory complex comprising PECAM-1, VEGFR2 and VE-cadherin and integrins at the cytoskeleton–extracellular matrix interface^{13,17} (Fig. 3a). We analysed the relationship between PLXND1 and these mechanical ‘hotspots’ in ECs. En face confocal imaging revealed robust and similar expression of PLXND1 in ECs in both the arch and descending aorta and colocalization with PECAM-1 at cell–cell junctions (Extended Data Fig. 7a). Staining was specific as it was not observed in *Plxnd1*^{IECKO} aortas (Extended Data Fig. 7b). SEMA3E was also observed in en face sections and expression of SEMA3E was found to be lower in the arch (Extended Data Fig. 7c). Co-immunoprecipitation experiments showed a flow-induced association of PLXND1 with components of the junctional mechanosensory complex (PECAM-1, VEGFR2, VE-cadherin and the p85 subunit of PI3K) (Extended Data Fig. 7d). To explore whether PLXND1 is just another component of the junctional complex or whether it operates upstream of the junctional complex, we used immunoprecipitation to analyse complex formation at the level of the junctional mechanosensory complex and integrin–matrix adhesions. We found that responses at the junctional complex, such as shear-stress-induced phosphorylation of VEGFR2 and association of the p85 subunit of PI3K and VE-cadherin with VEGFR2¹³, were all abrogated by knockdown of PLXND1 (Fig. 3b). Consistent with this observation, both the inhibition of the VEGFR2 receptor kinase (Fig. 3d) and the deletion of PECAM-1 abrogated force-induced signalling, suggesting that junctional mechanosensory components are necessary intermediates for the PLXND1-mediated force response (Extended Data Fig. 8a). Similarly, flow-induced complex formation at integrin–matrix adhesions (as assayed by association of Shc with integrin $\alpha_v\beta_3$)^{18,19} was also strongly reduced with loss of PLXND1 (Fig. 3c). A previous study has highlighted a role for PIEZO1-mediated and $G\alpha_q/G_{11}$ -mediated mechanosignalling, although there are conflicting reports as to whether these pathways are linked^{20,21} or independent of each other^{22,23}. Force application on PLXND1 showed that loss of $G\alpha_q/G_{11}$ abolished the PLXND1 force response, whereas knockdown of PIEZO1 had no effect (Extended Data Figs. 1e, f, 8b, c).

To further investigate the molecular mechanisms, we examined the role of the PLXND1 coreceptor neuropilin-1 (NRP1). NRP1 is a cell-surface transmembrane protein that acts as a SEMA3 and VEGF coreceptor for PLXND1 and VEGFR2, respectively²⁴, and its presence in neurons switches the SEMA3E signal from repulsion to attraction²⁵. We found that NRP1 was required for the PLXND1-mediated force response, as both knockdown (Extended Data Fig. 1d) and inhibition of NRP1 abolished the force-induced phosphorylation of vinculin (Fig. 3d). We also observed that shear stress induced the formation of a complex between PLXND1, VEGFR2 and NRP1 (Fig. 3e, f) and this complex was dependent on NRP1 (Fig. 3g). Taken together, these data show that PLXND1 associates with NRP1 and VEGFR2 in response to flow and operates upstream of both the junctional complex and integrins.

To test whether PLXND1 (and its molecular partners) is sufficient to confer mechanosensitivity in a heterologous cell line, we transfected COS-7 cells with plasmids expressing PLXND1, NRP1 and/or VEGFR2 and applied shear stress (Fig. 3h). These cells do not express any of the components of the junctional complex (that is, PECAM-1 or VE-cadherin) and are therefore an ideal system to monitor mechanical responses that are specifically due to PLXND1. COS-7 cells expressing all three proteins (VEGFR2, NRP1 and PLXND1) showed activation of early signalling responses, including phosphorylation of VEGFR2, association of VEGFR2 with Src tyrosine kinase, and PLXND1–VEGFR2–NRP1 complex formation in response to shear stress. Notably, none of these responses occurred in the absence of PLXND1, thus providing further evidence that PLXND1 is a specific and direct force sensor (Fig. 3i). Overall, these data

provide evidence that PLXND1 is necessary and sufficient for the shear-stress-induced response. To further demonstrate that PLXND1 operates as a specific force sensor, we applied force on other elements of the complex. As shown in Extended Data Fig. 9, application of force on either NRP1 or VEGFR2 did not elicit downstream responses. Taken together, these data show that PLXND1 is a specific and direct mechanosensor.

The mechanical response of PLXND1 is in stark contrast to the ligand response, as force on PLXND1 increases focal adhesions, whereas SEMA3E treatment reduces focal adhesions and leads to the collapse of the actin cytoskeleton⁴ (Extended Data Fig. 6b). Structure–function studies of semaphorins, plexins and their cognate complexes have established that the ligand-binding response requires a dimeric semaphorin to engage the N-terminal SEMA domains of two plexin receptors²⁶. Recent crystal structures and negative-stain electron-microscopy analyses of the entire, 10-domain class-A plexin (PLXNA) ectodomains revealed a distinctive ring-like conformation that is suitable for coupling extracellular semaphorin-based dimerization through to the transmembrane and cytoplasmic regions to transduce the ligand-binding response^{7,27}. However, the negative-stain electron-microscopy studies also revealed that the PLXNA ectodomain is capable of flexion, with distinctive minor populations of more-open conformations. We carried out negative-stain electron-microscopy analysis of the PLXND1 ectodomain and found evidence that it can flex to a more-open conformation, although the dominant state is ring-like (Fig. 4a, b and Extended Data Fig. 10a). We speculated that the ability to have flexion and switch between these two conformation states might provide an explanation for the binary nature of the functions of PLXND1 (Fig. 4c). To examine this, we generated the double mutant PLXND1(Y517C/A1135C), which is designed to promote the formation of an intramolecular disulfide bond between domain 1 and domain 9 of the PLXND1 ectodomain (Fig. 4d). On the basis of structural analyses, we predicted that the introduction of this disulfide bridge would lock the receptor ectodomain into the ring-like conformation, which would still enable the ligand-binding response by SEMA3E but would prevent the switch to the open and putative mechanosensory conformation. Purification of the protein and a subsequent quantitative assay using a thiol-reactive fluorescent dye, as well as negative-stain electron microscopy, demonstrated that the protein did indeed contain the desired covalent disulfide links (Extended Data Fig. 10b, c).

PLXND1-depleted ECs were infected with adenovirus expressing either wild-type or mutant PLXND1 and were assayed for their ability to respond to SEMA3E or mechanical force. Treatment with SEMA3E resulted in a decrease in focal adhesions in both wild-type and mutant PLXND1-expressing cells (Fig. 4e), showing that the PLXND1 ectodomain—when locked into a ring-like conformation—maintains its ability to bind to SEMA3E and signal to cause the disassembly of the cytoskeleton. We then tested whether trapping the PLXND1 in the semaphorin-binding ring-like conformation was permissive of its mechanosensory function. We found that cells expressing mutant PLXND1 did not respond to mechanical force, as assayed by the activation of early signalling responses (phosphorylation of VEGFR2, Akt and ERK1/2 in Fig. 4h), cytoskeleton signalling (phosphorylation of vinculin in Extended Data Fig. 11) and focal adhesion maturation (Fig. 4f). To further determine the requirement for PLXND1 flexion in mechanotransduction, we examined the effects of mutant PLXND1 in shear stress signalling. In contrast to ECs expressing wild-type PLXND1, ECs expressing mutant PLXND1 were unable to activate Akt, ERK1/2 or eNOS in response to shear stress (Fig. 4i). Additionally, reconstitution of mutant PLXND1 in COS-7 cells blocked early shear-stress responses, including phosphorylation of VEGFR2, association of VEGFR2 with Src tyrosine kinase and shear-stress-induced VEGFR2 and NRP1 complex formation (Fig. 4g). Taken together, these results demonstrate that

trapping PLXND1 in its ring-like conformation maintains its ligand-dependent signalling function but compromises its ability to sense and respond to mechanical force.

Our work identifies the semaphorin-binding receptor PLXND1 as a force detector in ECs. One of the best-characterized mechanosensors to date is the junctional mechanosensory complex, in which PECAM-1 is the molecule that can sense and respond to mechanical force^{12,13,28,29}. Given the proven crucial role of shear stress in cardiac and vascular development^{30,31}, it was always difficult to reconcile the lack of developmental defects in the PECAM-1 knockout mice. We now identify a previously undescribed mechanosensor in ECs that operates upstream of the junctional complex. We show that onset of shear stress induces the formation of a mechanocomplex of PLXND1–NRP1–VEGFR2; this complex requires the presence of NRP1 as well as flexion in the PLXND1 ectodomain. Endothelial PLXND1 regulates signals at junctions and integrins and downstream cellular responses to shear stress that ultimately regulate the site-specific distribution of atherosclerosis. The developmental cardiovascular defects observed in global³², as well as EC-specific³³, PLXND1 knockout mice are in agreement with a requirement for this mechanosensor during development, and—as our data now demonstrate—also in the adult. Despite the importance of mechanosensation in biology, knowledge of how mechanoreceptors detect physical force is limited. Our data identify a mechanosensor in ECs and provide a framework for understanding how ligand-dependent and mechanical signals can be channelled through a single receptor.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1979-4>.

- Davies, P. F. Flow-mediated endothelial mechanotransduction. *Physiol. Rev.* **75**, 519–560 (1995).
- Givens, C. & Tzima, E. Endothelial mechanosignaling: does one sensor fit all? *Antioxid. Redox Signal.* **25**, 373–388 (2016).
- Sakurai, A. et al. Semaphorin 3E initiates antiangiogenic signaling through plexin D1 by regulating Arf6 and R-Ras. *Mol. Cell. Biol.* **30**, 3086–3098 (2010).
- Aghajanian, H. et al. Semaphorin 3d and semaphorin 3e direct endothelial motility through distinct molecular signaling pathways. *J. Biol. Chem.* **289**, 17971–17979 (2014).
- Jongbloets, B. C. & Pasterkamp, R. J. Semaphorin signalling during development. *Development* **141**, 3292–3297 (2014).
- Hahn, C. & Schwartz, M. A. Mechanotransduction in vascular physiology and atherogenesis. *Nat. Rev. Mol. Cell Biol.* **10**, 53–62 (2009).
- Kong, Y. et al. Structural basis for plexin activation and regulation. *Neuron* **91**, 548–560 (2016).
- SenBanerjee, S. et al. KLF2 is a novel transcriptional regulator of endothelial proinflammatory activation. *J. Exp. Med.* **199**, 1305–1315 (2004).
- Hamik, A. et al. Kruppel-like factor 4 regulates endothelial inflammation. *J. Biol. Chem.* **282**, 13769–13779 (2007).
- Wu, C. et al. Mechanosensitive PPAP2B regulates endothelial responses to atherorelevant hemodynamic forces. *Circ. Res.* **117**, e41–e53 (2015).
- Zhang, S. H., Reddick, R. L., Piedrahita, J. A. & Maeda, N. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* **258**, 468–471 (1992).
- Collins, C. et al. Localized tensional forces on PECAM-1 elicit a global mechanotransduction response via the integrin–RhoA pathway. *Curr. Biol.* **22**, 2087–2094 (2012).
- Tzima, E. et al. A mechanosensory complex that mediates the endothelial cell response to fluid shear stress. *Nature* **437**, 426–431 (2005).
- Li, J. et al. Piezo1 integration of vascular architecture with physiological force. *Nature* **515**, 279–282 (2014).
- Xu, J. et al. GPR68 senses flow and is essential for vascular physiology. *Cell* **173**, 762–775 (2018).
- Bays, J. L., Campbell, H. K., Heidema, C., Sebbagh, M. & DeMali, K. A. Linking E-cadherin mechanotransduction to cell metabolism through force-mediated activation of AMPK. *Nat. Cell Biol.* **19**, 724–731 (2017).
- Katsumi, A., Orr, A. W., Tzima, E. & Schwartz, M. A. Integrins in mechanotransduction. *J. Biol. Chem.* **279**, 12001–12004 (2004).
- Jalali, S. et al. Integrin-mediated mechanotransduction requires its dynamic interaction with specific extracellular matrix (ECM) ligands. *Proc. Natl Acad. Sci. USA* **98**, 1042–1046 (2001).

19. Liu, Y., Sweet, D. T., Irani-Tehrani, M., Maeda, N. & Tzima, E. Shc coordinates signals from intercellular junctions and integrins to regulate flow-induced inflammation. *J. Cell Biol.* **182**, 185–196 (2008).
20. Albarrán-Juárez, J. et al. Piezo1 and G_q/G_{11} promote endothelial inflammation depending on flow pattern and integrin activation. *J. Exp. Med.* **215**, 2655–2672 (2018).
21. Wang, S. et al. Endothelial cation channel PIEZO1 controls blood pressure by mediating flow-induced ATP release. *J. Clin. Invest.* **126**, 4527–4536 (2016).
22. dela Paz, N. G., Melchior, B. & Frangos, J. A. Shear stress induces $G_{q/11}$ activation independently of G protein-coupled receptor activation in endothelial cells. *Am. J. Physiol. Cell Physiol.* **312**, C428–C437 (2017).
23. dela Paz, N. G. & Frangos, J. A. Rapid flow-induced activation of $G_{q/11}$ is independent of Piezo1 activation. *Am. J. Physiol. Cell Physiol.* **316**, C741–C752 (2019).
24. Gay, C. M., Zygmunt, T. & Torres-Vázquez, J. Diverse functions for the semaphorin receptor PlexinD1 in development and disease. *Dev. Biol.* **349**, 1–19 (2011).
25. Chauvet, S. et al. Gating of Sema3E/PlexinD1 signaling by neuropilin-1 switches axonal repulsion to attraction during brain development. *Neuron* **56**, 807–822 (2007).
26. Siebold, C. & Jones, E. Y. Structural insights into semaphorins and their receptors. *Semin. Cell Dev. Biol.* **24**, 139–145 (2013).
27. Suzuki, K. et al. Structure of the plexin ectodomain bound by semaphorin-mimicking antibodies. *PLoS ONE* **11**, e0156719 (2016).
28. Osawa, M., Masuda, M., Kusano, K. & Fujiwara, K. Evidence for a role of platelet endothelial cell adhesion molecule-1 in endothelial cell mechanosignal transduction: is it a mechanoresponsive molecule? *J. Cell Biol.* **158**, 773–785 (2002).
29. Collins, C. et al. Haemodynamic and extracellular matrix cues regulate the mechanical phenotype and stiffness of aortic endothelial cells. *Nat. Commun.* **5**, 3984 (2014).
30. Boselli, F., Freund, J. B. & Vermot, J. Blood flow mechanics in cardiovascular development. *Cell. Mol. Life Sci.* **72**, 2545–2559 (2015).
31. McCormick, M. E. & Tzima, E. Pulling on my heartstrings: mechanotransduction in cardiac development and function. *Curr. Opin. Hematol.* **23**, 235–242 (2016).
32. Gitler, A. D., Lu, M. M. & Epstein, J. A. PlexinD1 and semaphorin signaling are required in endothelial cells for cardiovascular development. *Dev. Cell* **7**, 107–116 (2004).
33. Zhang, Y. et al. *Tie2*cre-mediated inactivation of *plexinD1* results in congenital heart, vascular and skeletal defects. *Dev. Biol.* **325**, 82–93 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

Data reporting

No statistical methods were used to predetermine sample size. Mouse experiments were randomized.

Experimental mice

All mouse experiments were approved and authorized by both the University of Oxford Local Animals Ethics and Welfare Committee and by the UK Home Office. Project licences used in this work were 30/3080 and POC27F69A. *Plxnd1^{fl/fl}* mice were obtained from J. Epstein. To obtain the endothelial-cell-specific deletion of *Plxnd1*, *Plxnd1^{fl/fl}* mice were crossed with mice that expressed an inducible Cre recombinase under the *Cdh5^{cre}* driver, which were obtained from R. Adams. Three consecutive intraperitoneal injections of tamoxifen (2 mg each) in adult mice (6–8 weeks of age) resulted in the deletion of endothelial *Plxnd1* to generate PLXND1 inducible endothelial-cell knockout (*Plxnd1^{fl/fl}ECKO*) mice. For atherosclerosis studies, the *Plxnd1^{fl/fl}Cdh5^{cre}* mice were crossed into the hypercholesterolaemic apolipoprotein-E-deficient (*ApoE^{-/-}*) mouse background. Only female mice were used for atherosclerosis studies. All mice used in this study were maintained on a C57BL/6J background. For en face immunofluorescence analysis and qPCR of aortas, tissues were collected two weeks after the last tamoxifen injection. For atherosclerosis studies, a high-fat diet was commenced one week after the last tamoxifen injection.

All mice were housed in individually ventilated cages at 22 °C, with 56% relative humidity and a light–dark cycle of 12 h–12 h, and were fed a standard chow diet (B&K). For high-fat diet experiments carried out with hypercholesterolaemic *ApoE^{-/-}* mice, the mice were fed western RD (P) VP 25kGy diet containing 20% fat, 0.15% cholesterol (829108, SDS) for 10 weeks or 20 weeks. Water and food were available ad libitum at all times.

Genotyping

Genotyping was carried out using PCR analysis of DNA from ear notches, collected to identify the mice, using the Phire Tissue Kit (F140-WH, Thermo Scientific).

En face preparations

Mice were placed under a terminal general anaesthesia with isoflurane, followed by exsanguination and perfusion fixation with 4% paraformaldehyde. The entire length of the aorta was dissected out and the surrounding connective tissue and adventitial fat were removed. The aorta was fixed in 4% paraformaldehyde and stored at 4 °C in PBS until staining. Atheroprone areas from the inner curvature of the aortic arch were isolated and atheroprotective areas from the thoracic aorta were dissected and processed for immunofluorescence studies.

Oil-red-O staining for atherosclerosis studies

Fixed aortas were rinsed in absolute propylene glycol and stained with oil red O (O1516, Sigma Aldrich). After washing in 85% propylene glycol solution and distilled water, the aortas were opened longitudinally to the iliac bifurcation and a coverslip was placed to flatten down the aorta with the endothelial surface facing upwards. Images were acquired using an Olympus SZX7 fitted with a 1× lens and image processing was performed using Image-Pro (Media Cybernetics). The plaque area was quantified as a percentage of the area of both the total aorta and the aortic arch.

Lipid profile analysis

Blood was sampled by cardiac puncture under terminal general anaesthesia in plasma collection tubes. Plasma samples were shipped to MRC Harwell where they were analysed for total cholesterol, triglycerides, high-density lipoprotein and low-density lipoprotein levels on an automated AU680 Clinical Chemistry Analyser.

Cell culture, shear stress and transfections

Bovine aortic endothelial cells (BAECs), *Pecam1*-knockout (*Pecam1^{-/-}*) and PECAM reconstituted (*Pecam1^{+/+}*) mouse cells were cultured as previously described¹³. Mouse lung ECs were isolated from *Plxnd1^{fl/fl}* mice and maintained in EGM2 growth medium (Lonza), supplemented with 10% fetal bovine serum (FBS). COS-7 cells were maintained in DMEM with 10% FBS. All cell types were maintained at 37 °C in 5% CO₂ in a humidified incubator. Cells were subjected to shear stress using either a parallel plate chamber¹³ or a cone-and-plate viscometer as previously described¹⁰. For experiments using a SEMA3E-blocking antibody, cells were treated with the antibody for 1 h before and during shear stress. siRNA reverse transfections of *Plxnd1* and *PLXND1* in mouse ECs and BAECs were performed using the Lipofectamine RNAiMAX Reagent (Invitrogen). siRNAs used in this study were from Dharmacon and are described below.

Transfections of plasmids expressing NRP1, VEGFR2 and PLXND1 in COS-7 cells were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

siRNA knockdown in mouse ECs

The Acell mouse *Plxnd1* SMARTpool consisted of GUAUCGACCAC AGAUC AUG, CGUGGACCUUGAAUGGUU, CUAUUUAUAAACAGAUCCAA and CCAACAAGCUUCUGUACGC; The Acell mouse *Piezo1* SMARTpool consisted of CUAUCAGACACCAUUUAUC, GCCUCAUCCUCUAAUUAUGU, UCAUCAUCUCUAAGAAUUAU and CUGUUACGCUUCAAUGCUC; the Acell mouse *Gna11* SMARTpool comprised CCAUUUUCUAAAGU UAUUGA, CUUUUGAGCACCAGUAUGU, CUGUGACCCUUGUAUUAUA and CUGUCAGAUUUUCUUUACUU; the Acell mouse *Gnaq* SMARTpool comprised UUGUCAAGUUGUACGAAUU, CCAGGAUCAUAAGUGUUA, GUUAUGUGCAAUUAUGAAU and CGAUCAUACUAGGAGGGAU; the Acell human *NRP1* SMARTpool consisted of GCAGGAUUUUCCAUCGUU, CUUGAAUGCACUUUAUUAUG, UGGUUUAUCCUCAUUCUUUAU, UCCU GGAAUUUGAAAGCUU. The scramble siRNA was ON-TARGET plus non-targeting pool comprising UGGUUUACAUGUCGACUAA, UGGU UUAUCAUGUUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUU ACAUGUUUCCUA.

siRNA knockdown in BAECs

A *PLXND1* custom duplex comprising GGGAAACAUCGAGGCCAAUU and UUGGCCUCGAUGUUUCCCUU.

RNA extraction and qPCR

Total RNA extraction was performed from cells or from tissue using the RNeasy Plus Mini kit (Qiagen), with an additional genomic DNA-wipeout step. Reverse transcription was performed using the Superscript III cDNA synthesis kit. qPCR was performed in triplicate with SYBR green and a CFX96TM real-time system. Thermocycling conditions were 95 °C for 3 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 45 s. Gene expression was normalized to the constitutively expressed housekeeping gene 18S rRNA, and relative expression was calculated and plotted using the $\Delta\Delta C_t$ method. Primer sequences used were as follows. *Klf2*, 5'-CTAAAGGCGCATCTGCGTA-3', 5'-TAGTGGCGGGTAAGCTCGT-3'; *Klf4*, 5'-CGACTAACCCTGTGGCGTGA-3', 5'-GAGGTCTGTGAAT CCTCGG-3'; *Mcp1*, 5'-CATCCACGTGTGGCTCA-3', 5'-GATCATCT TGCTGGTGAATGAGT-3'; *Vcam1*, 5'-GCTATGAGGATGGAAGACTCTGG-3', 5'-ACTTGTGCAGCCACCTGAGATC-3'; 18S rRNA, 5'-AGGAATTGAC GGAAGGGCACCA-3', 5'-GTGCAGCCCCGACATCTAAG-3'.

Immunofluorescence

The permeabilization of tissues and cells was performed by incubation with 0.5% Triton X-100 overnight and 0.2% Triton X-100, respectively, and cells were blocked with 10% normal goat serum and 1% BSA. Inner curvatures of the aortic arch were incubated with primary antibodies (CD106 (VCAM-1, 553330, BD Biosciences) and MCP-1 (ab7202, Abcam))

and descending aorta segments were incubated with primary antibodies (PLXND1 (PA5-21605, Thermo Fisher Scientific) and PECAM-1 (553369, BD Biosciences)) before incubation with Alexa Fluor 488- and Alexa Fluor 568-conjugated secondary antibodies (1:100; Invitrogen). Cells were subjected to shear stress or tissues were incubated at 4 °C overnight in β -catenin (610153, BD Biosciences) followed by 1 h incubation with Alexa Fluor 488-conjugated phalloidin (Invitrogen) at room temperature and DAPI (Invitrogen). Tissues were mounted en face with Prolong Gold Antifade mountant (Invitrogen) for confocal imaging using an Olympus FluoView3000.

Image analysis

For the quantification of in vitro flow experiments, cell alignment in the direction of the flow was determined by measuring the angle between the flow direction and the long axis of the cell as determined visually³⁴. Cell elongation was estimated as the ratio of cell length to cell width in both in vitro and in vivo studies³⁵. Measurement of the fluorescence intensity of VCAM-1, MCP-1 and phalloidin was performed using ImageJ software (options: Analyze, Set measurements, Mean gray value, Measure). Quantification of the colocalization was performed using the colocal2 plugin in ImageJ.

Co-immunoprecipitation and western blotting

Cells were collected in lysis buffer as previously described¹⁹ and supplemented with protease and phosphatase inhibitor cocktail tablets. Lysates were precleared with 10 μ l protein A/G plus sepharose beads (Santa Cruz Biotechnology) for 1 h at 4 °C. The precleared lysates were then incubated with 20 μ l of protein A/G plus sepharose beads, which had previously been coupled with the appropriate primary antibody for 2 h at 4 °C on an orbital shaker. The beads were washed three times with lysis buffer supplemented with protease and phosphatase inhibitors. The immunoprecipitation complexes were eluted from the beads by boiling in 2 \times SDS buffer for 5 min.

For all western blotting analyses, protein lysates and co-immunoprecipitation complexes were resolved on a 4–12% gradient gel with the appropriate primary antibodies and IRDye-conjugated anti-mouse, anti-goat or anti-rabbit secondary antibodies, as appropriate. Images were acquired on a LICOR Odyssey infrared scanner. Densitometric quantification of bands was performed using the ImageStudio software (LICOR Biosciences).

Inhibitors, antibodies and other reagents

The antibodies used for western blotting included phosphorylated (p)-ERK1/2 (T202/Y204), total (t)-ERK1/2, p-Akt (S473), t-Akt, p-eNOS (S1177), p-VEGFR2 (Y1175), t-VEGFR2 (all antibodies from Cell Signaling Technology), t-eNOS (BD Biosciences), p-vinculin (Y822) (Abcam), t-vinculin (Sigma Aldrich), PI3K/p85 (Upstate), integrin $\alpha_3\beta_3$ (clone LM609, Merck), Shc (Abcam), VE-cadherin (Santa Cruz), PLXND1 (Thermo Fisher Scientific and Abcam), PIEZO1 (Abcam), G $\alpha_{q/11}$ (Santa Cruz Biotechnology) and Src (Upstate).

The inhibitors used in the study included the VEGFR2 tyrosine kinase inhibitor SU1498 (Sigma Aldrich). Recombinant SEMA3E was purchased from R&D Systems (Bio-technique) and used at 10 μ M. The NRP1-blocking antibody was purchased from R&D Systems and the SEMA3E-blocking antibody was from Thermo Fisher Scientific USA.

Bead pulling/magnetic tweezer system

Tosyl-activated paramagnetic beads (4.5 μ m) were washed with PBS and coated with an antibody against the extracellular domain of PLXND1 (Santa Cruz) or CD44 (clone 5D2-27 from the Developmental Studies Hybridoma Bank, USA). Beads were quenched in 0.2 M Tris, pH 7.4 before use to eliminate any remaining tosyl groups. ECs were incubated with the beads (and inhibitor or blocking antibody, if appropriate) before force application for 5–30 min at 37 °C. For immunofluorescence, ECs grown on fibronectin-coated coverslips were fixed for 20 min in PBS containing 2% formaldehyde, permeabilized with 0.2% Triton X-100 and blocked

with 10% goat serum for 1 h at room temperature. Antibody incubations for vinculin and HUTS4 were performed as previously described¹². Focal adhesion numbers were quantified as previously described³⁶. Ligated β 1 integrin staining was quantified by determining the mean fluorescence intensity using ImageJ software. To analyse the phosphorylation of vinculin, cells were lysed as described above and lysates were immunoblotted with a primary antibody against p-vinculin (Abcam).

Calcium imaging

BAECs were cultured in 33-mm glass-bottom dishes to form a sub-confluent monolayer. After the cells had fully attached and spread, 4 μ M of Fluo-8 AM, a calcium-binding dye (Abcam), was added to the medium. Cells were incubated for 30 min with the dye. Beads conjugated to either PLXND1 or poly-L-lysine were added to the cells and incubated for another 30 min. To assess the calcium influx as a result of mechanical stimulation, cells with Fluo-8 AM and magnetic beads were subjected to 1 nN force applied with magnetic tweezers. Time-lapse videos of epifluorescent calcium imaging were acquired with a Nikon Ti-e microscope (60 \times objective) during 10 s prestimulation, 20 s stimulation and 30 s poststimulation.

Acquired image sequences were analysed by measuring mean fluorescence intensity (mean pixel value) for each cell at each frame. Mean peak amplitude for each phase (prestimulation, poststimulation and during stimulation) was calculated and normalized to the prestimulation fluorescence intensity for each cell.

SEMA3E challenge

BAECs in which endogenous PLXND1 was knocked down with siRNA were infected with either wild-type or mutant PLXND1-expressing adenoviruses. Cells were serum-starved and treated with recombinant SEMA3E before processing for immunofluorescence with phalloidin, DAPI and anti-vinculin antibody. Images were taken on a Zeiss LSM 880 Airy Scan Confocal microscope and analysed using ImageJ³⁷ using an in-house-generated macro to measure the cell area, focal adhesion number and focal adhesion area. Statistical analyses were performed using GraphPad Prism 7. Comparisons between groups were assessed by two-way analysis of variance (ANOVA) with a Tukey multiple-comparisons post hoc test. Difference were considered significant when $P < 0.05$.

Site-directed mutagenesis

To lock the ectodomain of PLXND1 in the ring-like conformation, we designed a double mutant by introducing two single point mutations, Y517C and A1135C, in the SEMA domain (domain 1) and IPT5 domain (domain 9), respectively. Site-directed mutagenesis of full-length PLXND1, and of the PLXND1 ectodomain, was carried out by multiple-step overlap-extension PCR, and the resulting PCR products were cloned into a pHLSec vector³⁸.

Protein production

Constructs encoding the ectodomain (residues 47–1271) of mouse PLXND1 or double-mutant PLXND1(Y517C/A1135C) were cloned into the pHLSec vector in frame with a C-terminal hexahistidine (6His) tag. Protein was produced by transient transfection in HEK293T cells at 37 °C. Conditioned medium was collected five days after transfection and buffer was exchanged using a QuixStand diafiltration system (GE Healthcare). The double mutant of PLXND1 was secreted at a similar level as the wild-type protein. Proteins were purified by immobilized metal-affinity chromatography using a HisTrap FF column (GE Healthcare) followed by size-exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Healthcare).

Alexa Fluor labelling of PLXND1(Y517C/A1135C) for validation of disulfide-bond formation

PLXND1(Y517C/A1135C) at a concentration of 10 μ M in PBS was labelled with a 20-fold molar excess of a thiol-reactive fluorescent dye, Alexa

Article

Fluor 488 C5 maleimide (Thermo Fisher Scientific), and the reaction was allowed to proceed for 24 h at 6 °C in the dark. Unreacted dye was removed from the labelled protein using a Sephadex G-25 column (GE Healthcare). The degree of labelling (n) was determined using Eq. (1):

$$n = \frac{A_{488} M}{\epsilon c} \quad (1)$$

in which A_{488} is the absorbance at 488 nm, M is the molecular mass of the protein, ϵ is the molar extinction coefficient of the dye and c is the protein concentration in mg ml^{-1} . Hen egg ovalbumin (GE Healthcare) was used as a positive control.

Negative-stain electron microscopy

A drop of 2.5 μl freshly gel-filtrated PLXND1 ectodomain at a concentration of 1–5 $\mu\text{g ml}^{-1}$ in 10 mM HEPES, pH 7.5 and 150 mM sodium chloride was adsorbed to a newly glow-discharged carbon-coated copper grid, washed with two drops of 50 μl deionized water, and stained with two drops of 50 μl 0.75% uranyl formate. The excess stain on the grids was removed with filter paper before air-drying. Samples were imaged at room temperature using an FEI Tecnai T12 electron microscope equipped with a LaB6 filament operating at an acceleration voltage of 120 kV and a dose of 15 electrons per \AA^2 . Images were taken using a 4,000 \times 4,000 FEI Eagle TM CCD camera at a magnification of 57,000 \times with under-focus values ranging from 1.0 to 1.5 μm and a pixel size of 2.16 \AA . The particle images were normalized, rescaled, filtered before being subjected to reference-free classification in EMAN2³⁹. The PLXND1 structural models were generated manually using The PyMOL Molecular Graphics System (Schrödinger).

Cloning and adenoviral generation

Wild-type and mutant PLXND1 were cloned into the pENTR/TOPO entry vector of the Gateway System (Invitrogen) using the KOD Hot Start High Fidelity polymerase. After confirmation of successful cloning by Sanger sequencing, the constructs were sub-cloned into the pAd/CMV/V5-Dest destination vector by LR Clonase II reaction. All steps were performed according to the manufacturer's instructions. The destination vector was linearized by PacI digestion and transfected into HEK293A cells for adenoviral generation and subsequent amplification according to the manufacturer's instructions. In experiments in which adenoviral overexpression was used, endogenous levels of PLXND1 were knocked down with the siRNA pool to minimize any background signals.

Statistics

Data are mean \pm s.e.m. All experiments were performed at least three times independently. Statistical significance was tested using either an ANOVA or unpaired Student's t -tests. Data were tested for normality using the Shapiro–Wilk test and equality of variance using the Levene test. Where necessary, data were log-transformed before being analysed for statistical significance. All image analysis was performed by operators who were blinded to the treatments administered.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The datasets generated during and/or analysed during this study are either included within the manuscript or are available from the corresponding author on reasonable request. Source Data for Figs. 1–4 and Extended Data Figs. 2–11 are provided with the paper. Gel source data can be found in Supplementary Fig. 1.

34. Baeyens, N. et al. Syndecan 4 is required for endothelial alignment in flow and atheroprotective signaling. *Proc. Natl Acad. Sci. USA* **111**, 17308–17313 (2014).
35. Sabine, A. et al. Mechanotransduction, PROX1, and FOXC2 cooperate to control connexin37 and calcineurin during lymphatic-valve formation. *Dev. Cell* **22**, 430–445 (2012).
36. Horzum, U., Ozdil, B. & Pesen-Okvur, D. Step-by-step quantitative analysis of focal adhesions. *MethodsX* **1**, 56–59 (2014).
37. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
38. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D* **62**, 1243–1250 (2006).
39. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).

Acknowledgements This work was supported in part by grants from the Wellcome Trust (Senior Research Fellowship to E.T.), BHF (PG/16/29/32128 to E.T.), John Fell Fund (to E.T.) the BHF Centre of Excellence, Oxford (RE/13/1/30181), Cancer Research UK and the UK Medical Research Council (C375/A17721 and MR/M000141/1 to E.Y.J.), and Wellcome Trust grant 203141/Z/16/Z supporting the Wellcome Centre for Human Genetics and MICRON imaging facility (<http://micronoxford.com>, supported by Wellcome Strategic Awards 091911/B/10/Z and 107457/Z/15/Z). We thank K. Channon and G. Douglas for providing the Home Office Project Licences under which part of the animal studies were performed, A. Jefferson for help with confocal imaging, for technical advice and access to equipment, V. Jain for providing the NRPI plasmid and L. Payne for help with qPCR.

Author contributions V.M. performed or was involved in most of the experiments and analyses. K.-L.P. performed en face staining and imaging of all aortas, staining and imaging for in vitro alignment, most qPCR experiments and quantification of the data. D.R. designed and validated the ring-locked PLXND1 mutant. K.N. performed activation of signalling mediators in response to shear stress and initial magnetic force application experiments. A.K. performed semaphorin challenge experiments and analysed the data. D.L. performed the calcium-imaging experiments. Y.K. provided structural analysis of the PLXND1 ectodomain. D.K. and M.A. carried out the negative-stain electron-microscopy analysis. J.H. performed the initial PLXND1 siRNA experiments. Y.F. provided the design of the cone-and-plate system. A.d.R.H. led and supervised the calcium-imaging experiments. J.S.R. cosupervised and interpreted data, conceived and developed the idea of the binary conformations of PLXND1 and performed cloning of the PLXND1 wild-type and mutant constructs into adenovirus. E.Y.J. led and supervised the structural-biology-based components of the study. E.T. initiated the project, generated research funds and ideas, directed and coordinated the project. V.M., J.S.R., E.Y.J. and E.T. designed experiments, interpreted data and wrote the manuscript, with inputs from all authors.

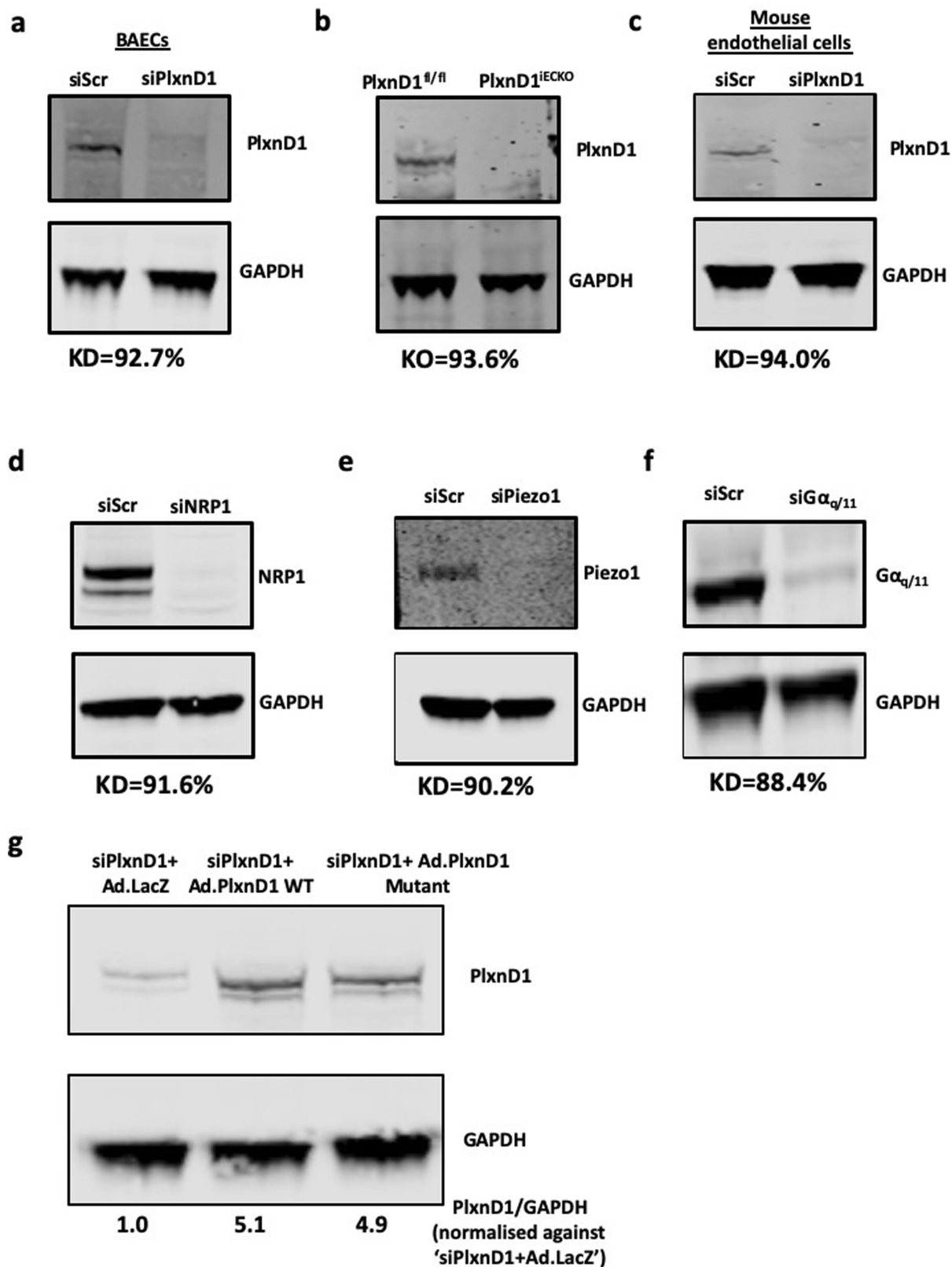
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1979-4>.

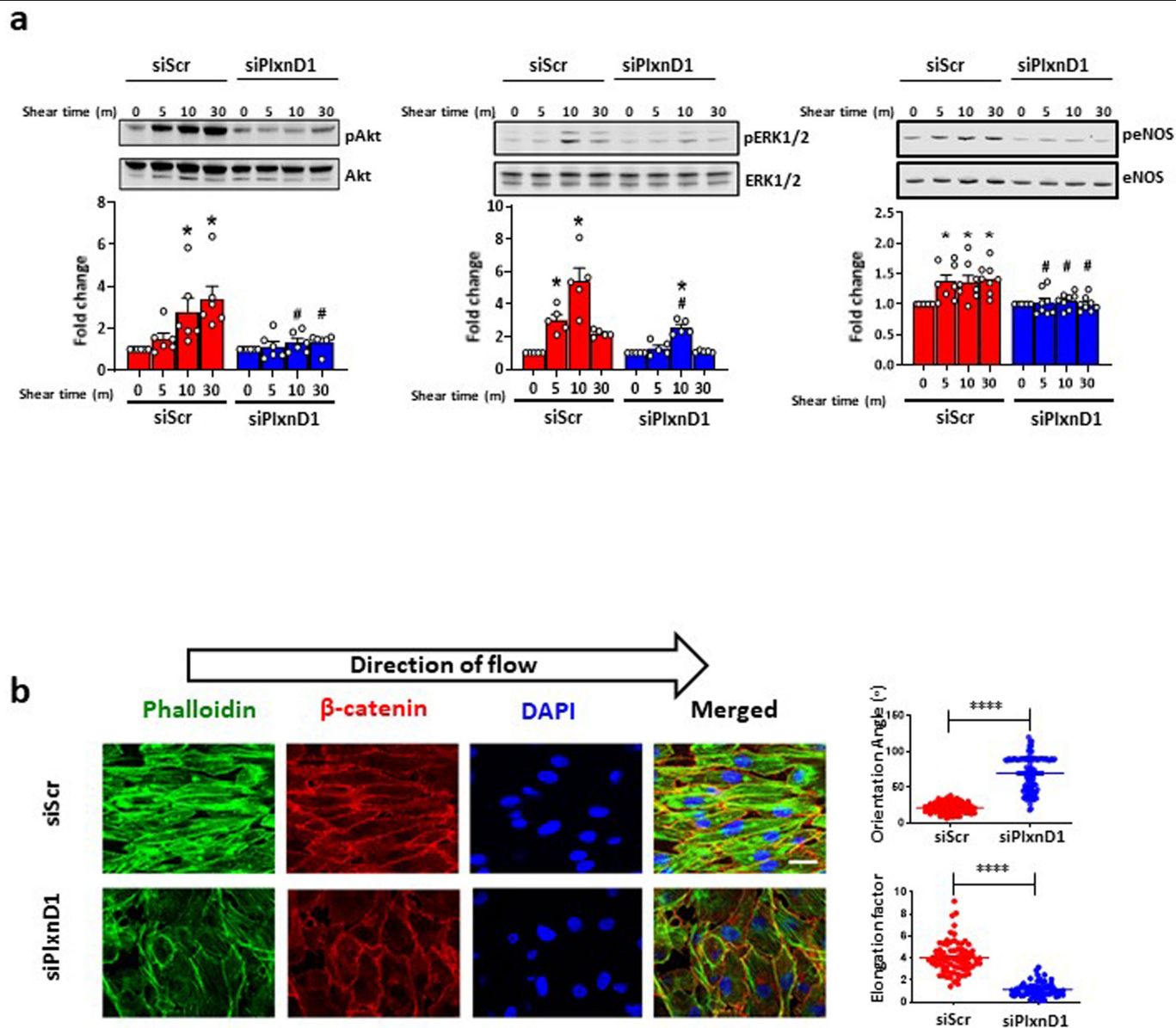
Correspondence and requests for materials should be addressed to E.T.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



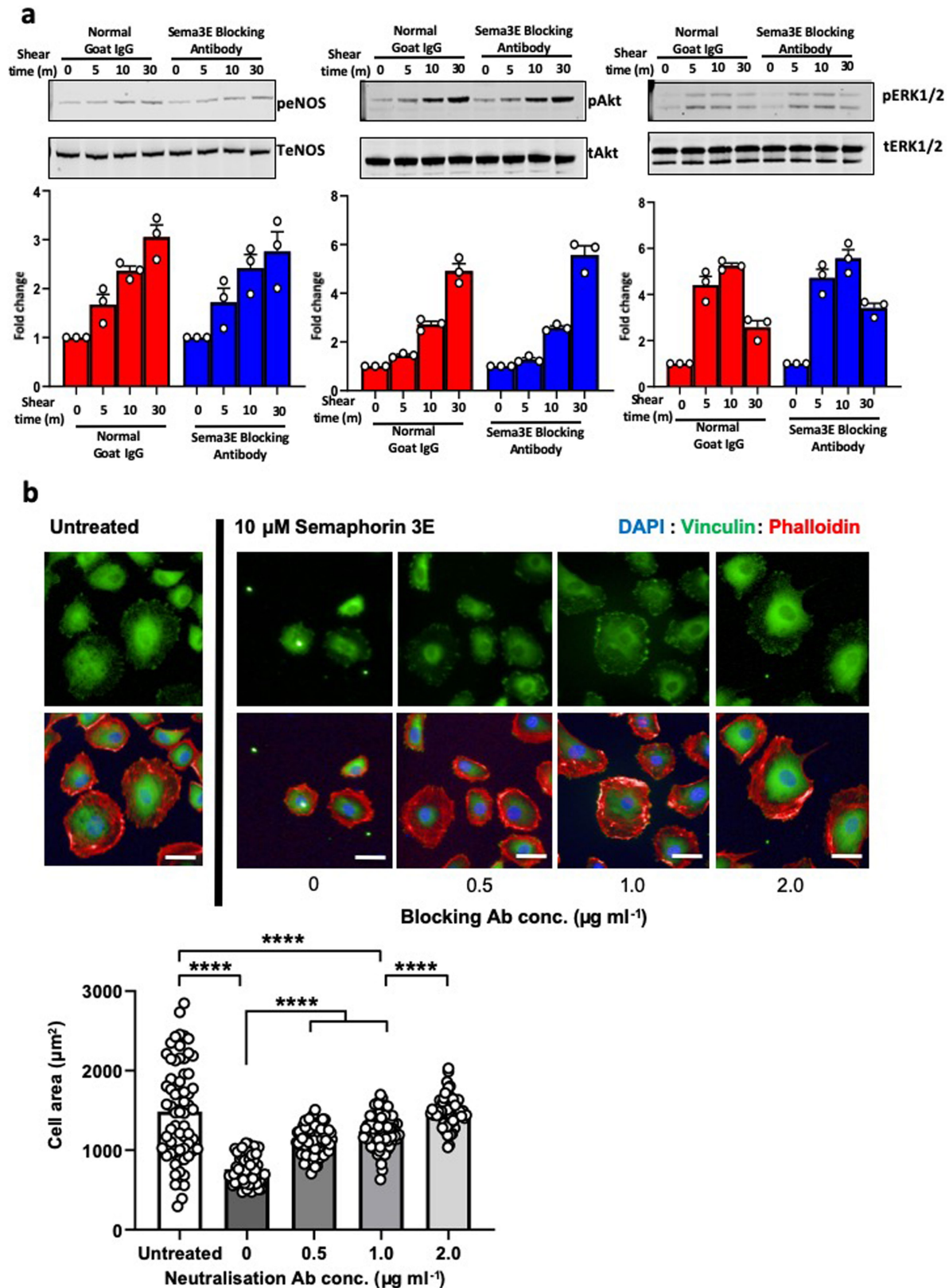
Extended Data Fig. 1 | Knockout or knockdown of PLXND1 and other genes in ECs. **a–f**, ECs were either isolated from *Plxnd1*^{fl/fl} and *Plxnd1*^{IECKO} mice, or treated with siRNAs to knockdown PLXND1, NR1, PIEZO1 and Gα_{q/11}. Knockdowns and knockouts were confirmed by western blotting, using GAPDH as a loading control. **g**, PLXND1 was knocked down in mouse ECs using a pool of siRNAs,

followed by infection with an adenovirus expressing either β-galactosidase (LacZ), or wild-type or mutant PLXND1. Protein levels were normalized to GAPDH. KD, mean knockdown efficiency based on $n = 3$; KO, mean knockout efficiency based on $n = 3$.



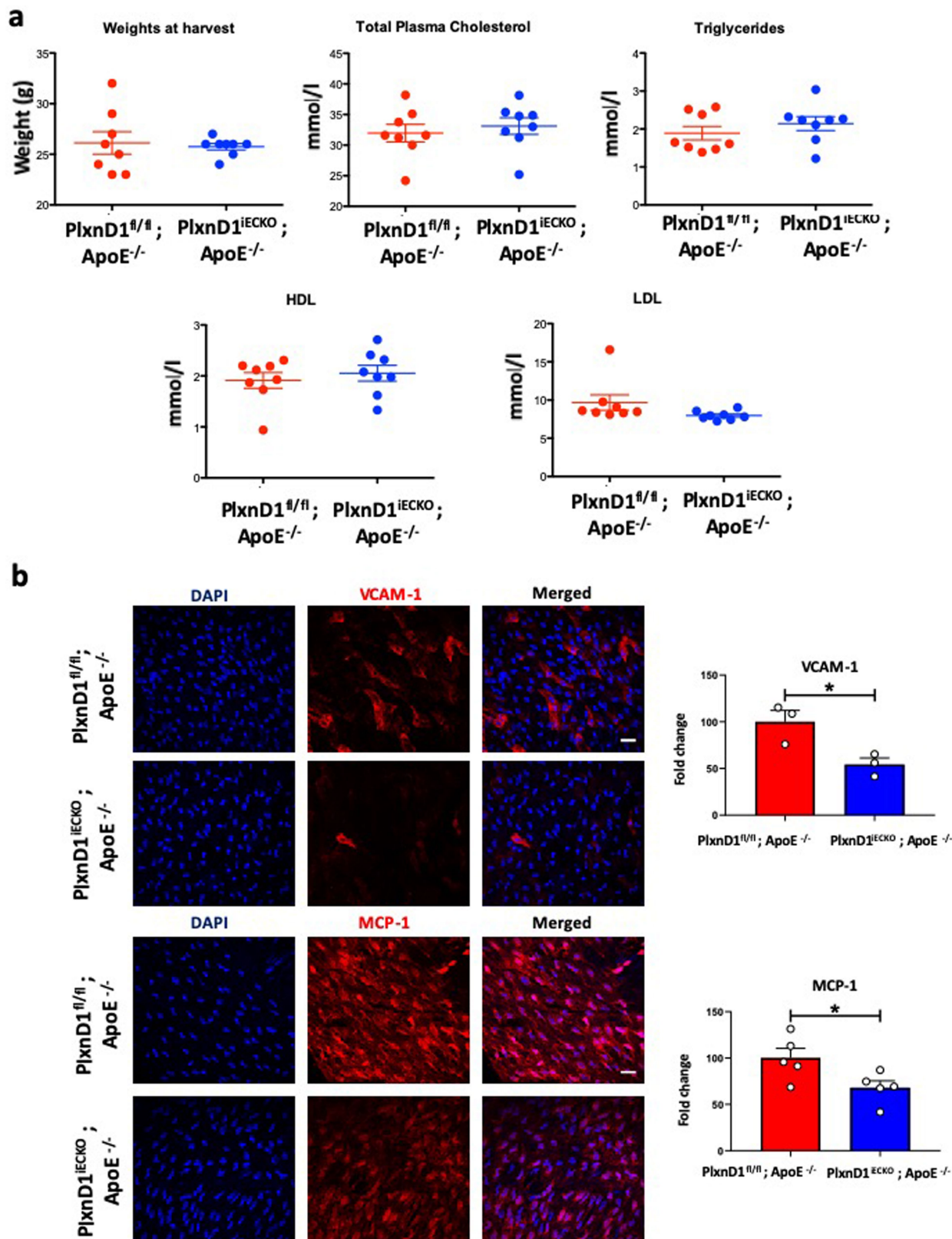
Extended Data Fig. 2 | PLXND1 mediates the EC response to fluid shear stress. **a**, BAECs were transfected with scrambled or *PLXND1* siRNA and exposed to laminar fluid shear stress (12 dynes cm^{-2}) using a parallel plate system for the indicated time periods. Phosphorylation of Akt ($n=6$), ERK1/2 ($n=5$) and eNOS ($n=8$) was determined by western blotting and quantified using Image Studio Lite v.5.2. Data are mean \pm s.e.m. P values were obtained by two-tailed Student's t -tests using GraphPad Prism. * $P < 0.05$ relative to the static condition; # $P < 0.05$ relative to the shear time point of the respective scrambled

siRNA. **b**, BAECs were transfected with scrambled or *PLXND1* siRNA and exposed to atheroprotective shear stress for 24 h. Cells were fixed and stained with phalloidin, DAPI and anti- β -catenin antibodies to visualize actin stress fibres, nuclei and cell junctions, respectively. Quantification of alignment was performed using ImageJ; $n > 50$ cells across 4 biological replicates (exact sample numbers are provided in the Source Data). Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -tests using GraphPad Prism. **** $P < 0.0001$.



Extended Data Fig. 3 | Mechanotransduction by PLXND1 is independent of its ligand-binding functions. a, BAECs were treated with SEMA3E-blocking antibody or control antibody ($1 \mu\text{g ml}^{-1}$) and exposed to fluid shear stress for the indicated times. Phosphorylation of eNOS, Akt and ERK1/2 was determined by western blotting and quantified using Image Studio Lite v.5.2. $n = 3$ biological repeats. Data are mean \pm s.e.m. **b,** BAECs were treated with SEMA3E-blocking antibody or control antibody for 1 h before exposure to SEMA3E for 30 min at the indicated concentrations. Cells were fixed and probed with anti-

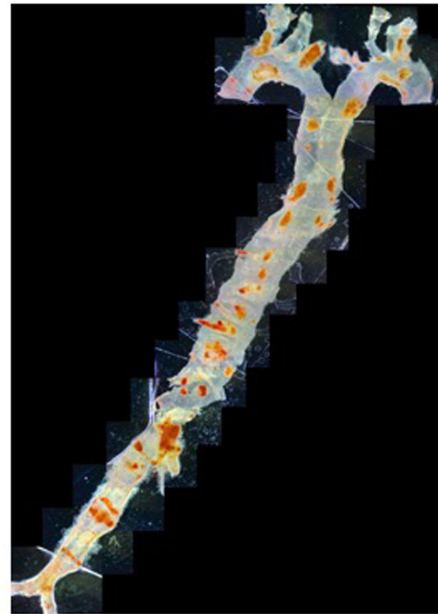
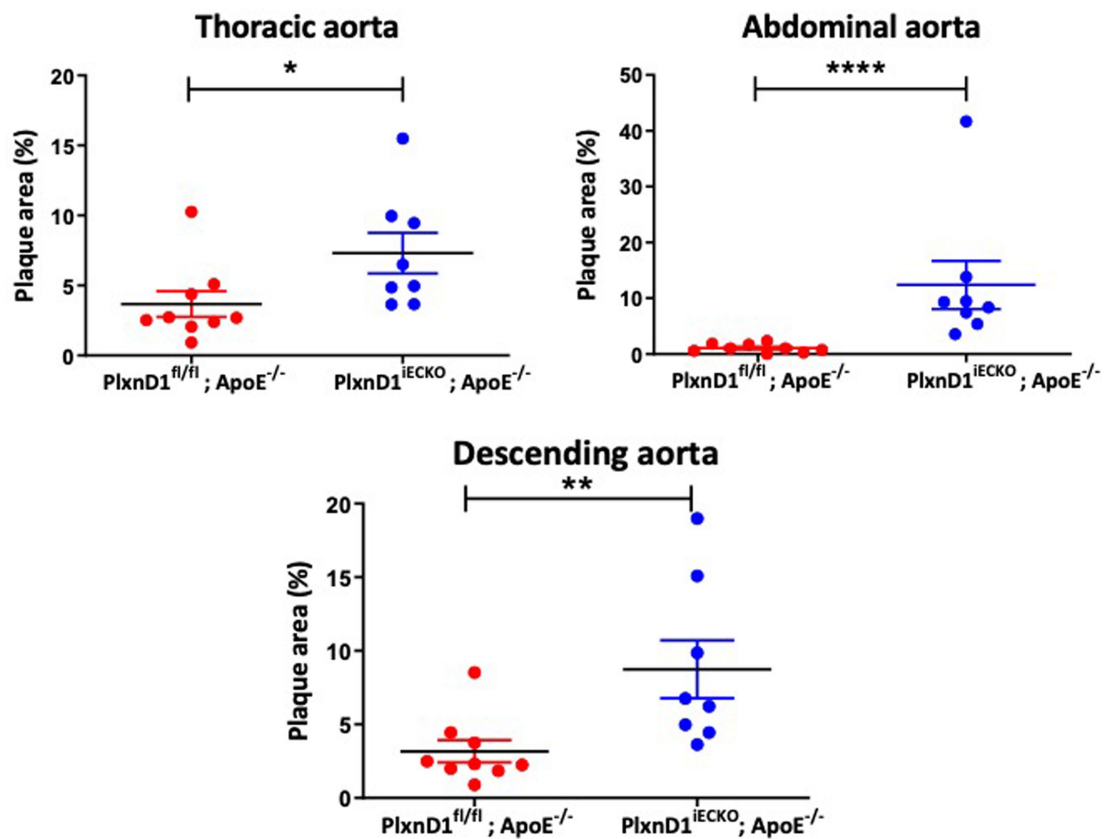
vinculin antibody, then stained with phalloidin and DAPI to visualize focal adhesions, actin stress fibres and nuclei, respectively. EC collapse was quantified by measuring the cell area using ImageJ. Data are mean \pm s.e.m. Significance was determined by ANOVA with a Tukey post hoc test using GraphPad Prism. **** $P < 0.0001$. $n = 59-82$ cells across 3 independent experiments (exact sample numbers are provided in the Source Data). Scale bar, $50 \mu\text{m}$.



Extended Data Fig. 4 | Lipid-profile analysis and expression of inflammatory markers in the aortic arch. a, Body weights and lipid-profile analysis of *PlxnD1^{fl/fl}; ApoE^{-/-}* and *PlxnD1^{IECKO}; ApoE^{-/-}* mice after 10 weeks of high-fat diet feeding (analysed at 16–17 weeks of age); $n = 8$. Data are mean \pm s.e.m.

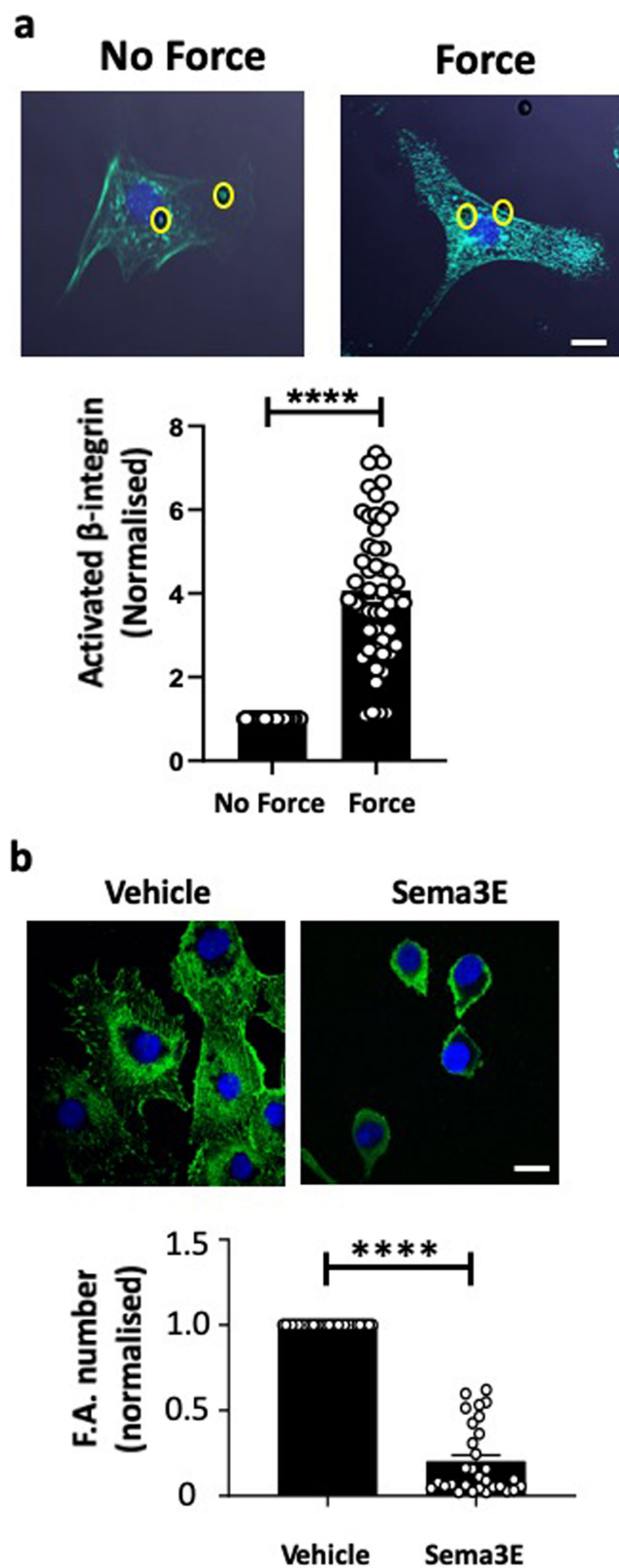
b, Representative en face preparations of aortic arches immunostained for

VCAM-1 ($n = 3$) and MCP-1 ($n = 5$) from *PlxnD1^{fl/fl}; ApoE^{-/-}* and *PlxnD1^{IECKO}; ApoE^{-/-}* mice with quantification of fluorescence intensity in fold change; 3–5 images were taken from the inner curvature of aortic arch of each mouse. Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -tests using GraphPad Prism. * $P < 0.05$.

a**PlxnD1^{fl/fl}; ApoE^{-/-}****PlxnD1^{IECKO}; ApoE^{-/-}****b****Extended Data Fig. 5 | Atherosclerosis in the descending aorta.**

a, Representative en face preparations of the whole aorta showing atherosclerosis in *PlxnD1^{fl/fl}ApoE^{-/-}* and *PlxnD1^{IECKO}ApoE^{-/-}* mice after 20 weeks of high-fat diet feeding, visualized by oil-red-O staining. **b**, Quantification of lesion area in the thoracic aortas, abdominal aortas and whole descending

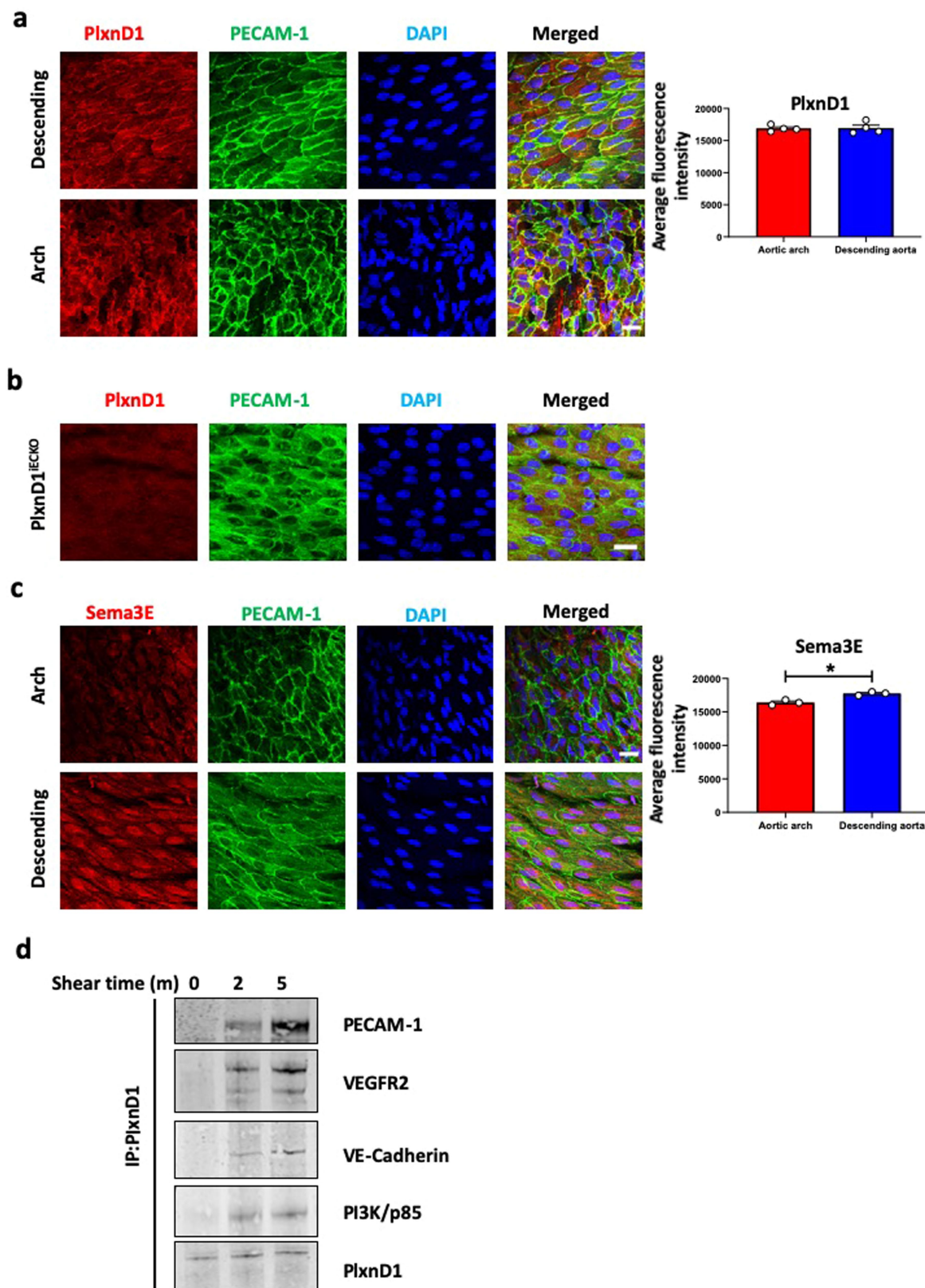
aortas (thoracic aorta and abdominal aorta) of $n = 9$ *PlxnD1^{fl/fl}ApoE^{-/-}* and $n = 8$ *PlxnD1^{IECKO}ApoE^{-/-}* mice. Data are mean \pm s.e.m. *P* values were obtained using two-tailed Student's *t*-tests using GraphPad Prism. **P* < 0.05, ***P* < 0.01, *****P* < 0.0001.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Mechanical force on PLXND1 results in integrin activation, whereas ligand stimulation causes ECs to collapse. a, BAECs were incubated with anti-PLXND1-coated beads and subjected to force (10 pN) for 5 min. ECs were fixed and stained with HUTS4 antibody to mark ligated $\beta 1$ integrin. Mean fluorescence intensity was quantified using ImageJ software. Values were normalized to the no force condition. Locations of the beads are highlighted in yellow circles. $n = 50$ cells per condition from 3 independent experiments. Data are mean \pm s.e.m. P values were obtained using two-tailed

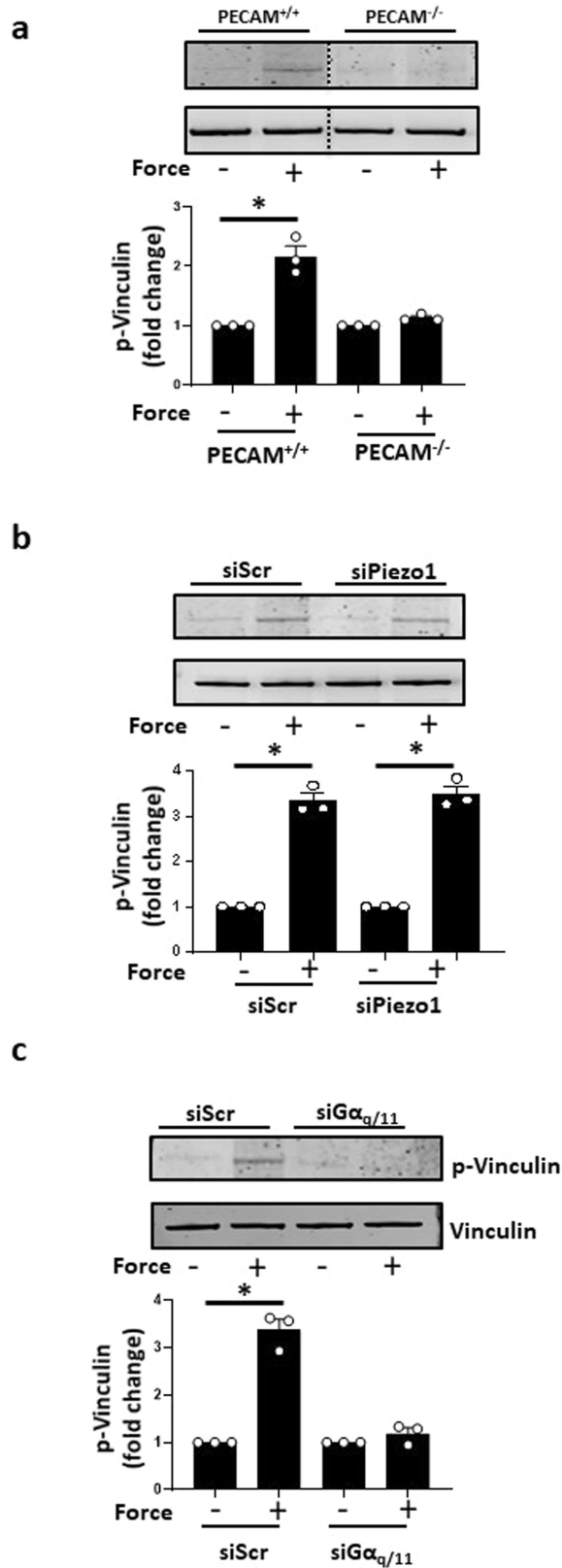
Student's t -tests using GraphPad Prism. **** $P < 0.0001$. Scale bar, 10 μm . **b,** BAECs were incubated with SEMA3E or vehicle, fixed and stained with anti-vinculin antibody to mark focal adhesions. Focal adhesion number was quantified using ImageJ software. Values were normalized to the vehicle condition. $n = 30$ cells per condition from 3 independent experiments. Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -tests using GraphPad Prism. **** $P < 0.0001$. Scale bar, 10 μm .



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | PLXND1 colocalizes and associates with members of the junctional mechanosensory complex, and its levels are not regulated by flow, in contrast to SEMA3E. **a**, The descending thoracic aorta or the inner curvature of aortic arches were isolated and prepared en face from wild-type mice and stained for PLXND1, PECAM-1 and DAPI. Quantification of PLXND1 levels was performed by fluorescence intensity measurement using ImageJ; 4–6 images were taken of tissue collected from $n = 4$ mice. Data are mean \pm s.e.m. Scale bar, 20 μ m. **b**, The descending thoracic aorta was isolated and prepared en face from *Plxnd1*^{IECKO} mice and stained for PLXND1 expression to assess the specificity of the PLXND1 immunostain. $n = 3$ mice all showed similar results. **c**, The descending thoracic aorta or the inner curvature of aortic

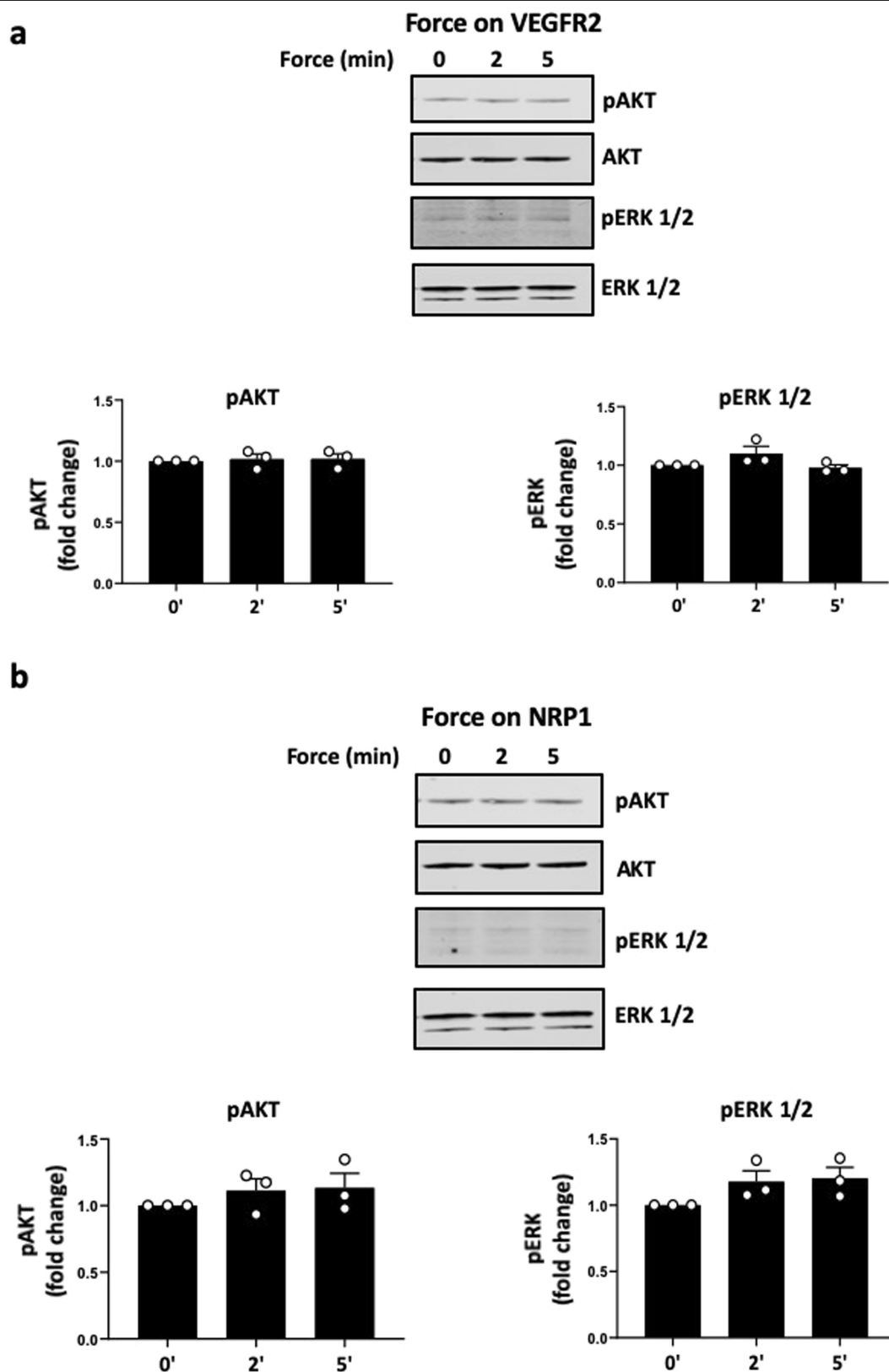
arches were isolated and prepared en face from wild-type mice and stained for SEMA3E and PECAM-1 expression and with DAPI. Quantification of SEMA3E levels was performed by fluorescence intensity measurement using ImageJ; 4–6 images were taken using tissue collected from $n = 3$ mice. Data are mean \pm s.e.m. *P* values were obtained using two-tailed Student's *t*-test using GraphPad Prism. $*P < 0.05$. Scale bar, 20 μ m. **d**, Mouse ECs were exposed to shear stress for the indicated times or left as static controls before immunoprecipitating PLXND1 and analysing its association with the junctional mechanosensory complex (PECAM, VE-cadherin and VEGFR2) as well as PI3K/p85. $n = 3$ independent experiments.



Extended Data Fig. 8 | See next page for caption.

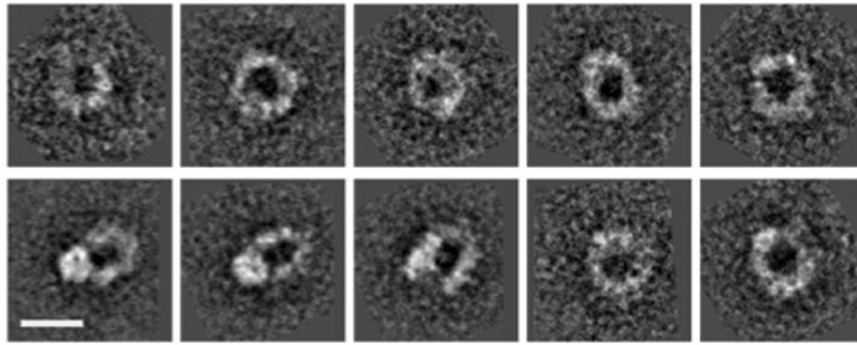
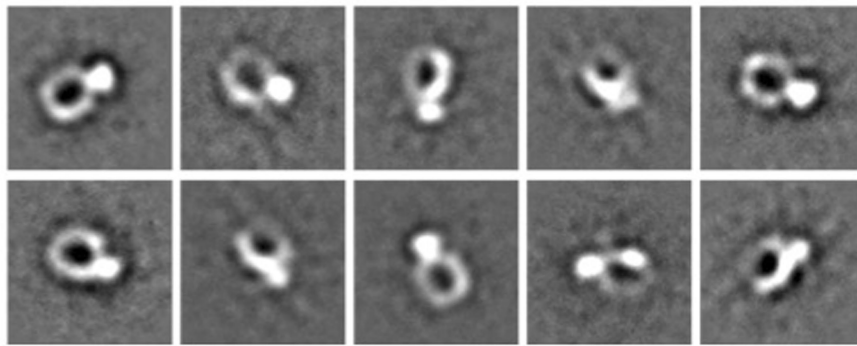
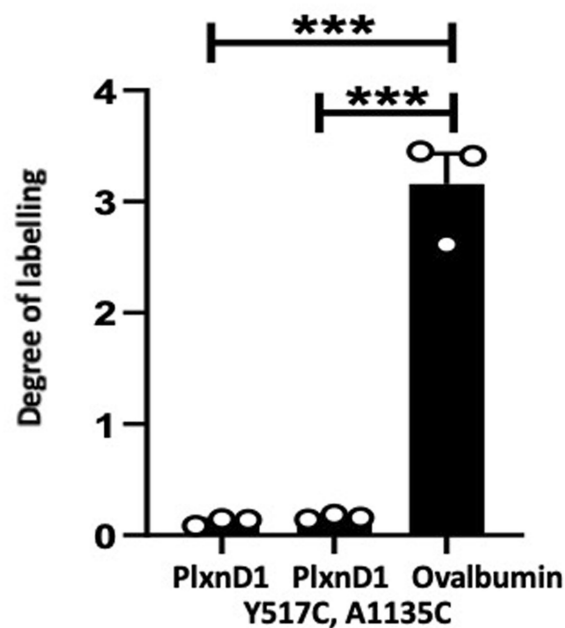
Extended Data Fig. 8 | Relationship between PLXND1 and other established mechanosensors. a, *Pecam1*^{+/+} and *Pecam1*^{-/-} ECs were incubated with anti-PLXND1-coated beads and subjected to force application for 5 min before the analysis of the phosphorylation of vinculin. *n* = 3 independent experiments. **P* < 0.05. **b, c,** Mouse ECs were treated with siRNAs against Piezo1 and Gα_{q/11},

incubated with anti-PLXND1-coated beads and subjected to force application for 5 min before the analysis of the phosphorylation of vinculin. *n* = 3. **P* < 0.05. Data are mean ± s.e.m. *P* values were obtained using two-tailed Student's *t*-tests using GraphPad Prism.



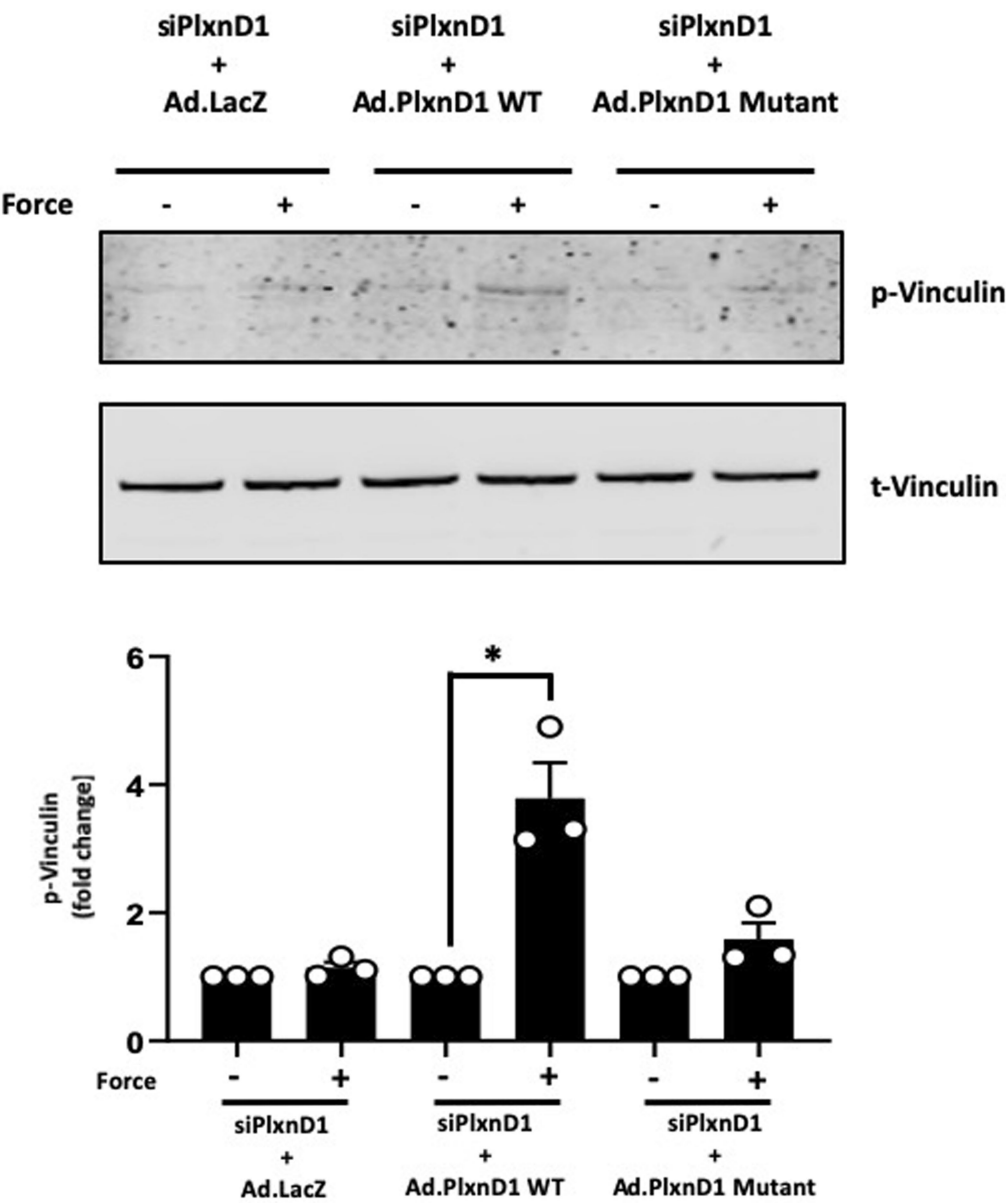
Extended Data Fig. 9 | Force application on other members of the PLXND1 mechanocomplex does not elicit a mechanotransduction response.
a, b, Mouse ECs were incubated with anti-VEGFR2 (**a**) or anti-NRP1 (**b**) antibody-coated beads and subjected to force (10 pN) for the indicated time periods.

Phosphorylation of Akt and ERK1/2 was determined by western blotting and quantified using Image Studio Lite v.5.2. $n = 3$ biological repeats. Data are mean \pm s.e.m.

a**b****c**

Extended Data Fig. 10 | Validation of the PLXND1 mutant. a, Negative-stain two-dimensional class averages of wild-type PLXND1 were obtained by classifying 1,305 particles into 10 classes. Scale bar, 10 nm. **b,** Negative-stain two-dimensional class averages of mutant PLXND1 were obtained by classifying 1,357 particles into 10 classes. **c,** The double mutant of PLXND1 was labelled with a thiol-reactive fluorescent dye, Alexa Fluor 488 C5 maleimide. The degree of labelling shows that the vast majority of PLXND1 mutant

molecules form the disulfide-linked bond and thus the ring of the majority of PLXND1-mutant molecules appears to be locked by the covalent bond. The degree of labelling for the hen egg ovalbumin, which we used as a positive control, is close to the number of free cysteines in ovalbumin. $n=3$ independent experiments. Data are mean \pm s.e.m. P values were calculated by two-tailed Student's t -tests using GraphPad Prism. *** $P < 0.001$.



Extended Data Fig. 11 | An open conformation of PLXND1 is required for force-dependent signalling. Mouse lung ECs in which endogenous PLXND1 was knocked down were infected with adenoviruses expressing β -galactosidase (Ad.LacZ), wild-type or mutant PLXND1 and incubated with

anti-PLXND1 paramagnetic beads, followed by force application for 5 min before lysing and assaying vinculin phosphorylation by western blotting. $n = 3$ biological repeats. Data are mean \pm s.e.m. P values were obtained using two-tailed Student's t -tests using GraphPad Prism. * $P < 0.05$.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Image Studio Lite Version 5.2.5 (LI-COR) was used for western blots, ZEN 2.3 SP1 (Zeiss Software) was used for microscopy image acquisition, EMAN2, Imagic,

Data analysis

Image Studio Lite Version 5.2.5 (LI-COR) was used for all western blotting analysis, EMAN2, Imagic, Pymol Molecular Graphics System Version 1.8.6.2 (Schrodinger Inc.), FIJI (Image J, NIH) or ImageJ Pro were used for analysis of microscopy data, GraphPad Prism 6 was used for statistical analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data for quantifications are available from the corresponding author upon request. There are no restrictions on data availability.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To determine the number of animals required to complete the proposed studies, we performed power analyses based on literature review and data from our laboratory using the SPSS statistical software. Too high numbers were avoided to keep the number of experimental animals as low as possible. All cell-based experiments had at least 3 biological replicates.
Data exclusions	No samples or animals were excluded.
Replication	Experimental findings were successfully replicated, by either the same author(s) or a different author.
Randomization	Age/sex matched litter mates were randomly assigned to groups. For magnetic force application experiments experiments, cells with 1-3 beads were randomly selected.
Blinding	The investigators were blinded to group allocation/treatments during data collection. Most of the analysis was performed in an automated/semi-automated fashion.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials Mouse cell lines, adenoviral constructs unique to this study can be obtained from the corresponding author upon reasonable request.

Antibodies

Antibodies used phospho(p)-ERK1/2T-202;Y-204 (9106, Cell Signaling Technology, 1:1000), total(t)-ERK1/2 (9102, Cell Signaling Technology,

1:1000), pAktS473 (4060, Cell Signaling Technology, 1:1000), tAkt (4691, Cell Signaling Technology, 1:1000), p-eNOSS1177 (9571, Cell Signaling Technology, 1:1000), pVEGFR2Y1175 (2478, Cell Signaling Technology, 1:1000), tVEGFR2 (2479, Cell Signaling Technology, 1:1000), t-eNOS (610296, BD Biosciences, 1:1000), p-VinculinY822 (ab200825, Abcam, 1:1000), t-Vinculin (V9131, Sigma Aldrich, 1:1000), PI3K/P85 (06195, Upstate, 1:1000), integrin $\alpha\beta$ 3 (clone LM609, Merck, 1:1000), Shc (ab15039, Abcam, 1:1000), VE-cadherin (C-19, sc-6458, Santa Cruz Biotechnology, 1:1000), Fibronectin (HFN7.1, Developmental Society Hybridoma Bank, 1:50), MCP1 (ab7207, Abcam, 1:50), beta-catenin (610153, BD Transduction Laboratories, 1:70), Alexafluor 488 goat anti-mouse (A11001, Invitrogen, 1:150), Alexafluor 568 goat anti-mouse (A11061, Invitrogen, 1:150), Alexafluor 568 goat anti-rabbit (A11011, Invitrogen, 1:150), Alexafluor 790 goat anti-rabbit (A11367, Invitrogen, 1:10000), Alexafluor 680 goat anti-rabbit (A21076, Invitrogen, 1:10000), Alexafluor 790 goat anti-mouse (A11375, Invitrogen, 1:10000), Alexafluor 680 goat anti-mouse (A21058, Invitrogen, 1:10000), Plexin D1 (E13, sc-46245, Santa Cruz Biotechnology), Plexin D1 (PA5-47012, ThermoFisher Scientific)

Validation

Antibodies were validated by the respective suppliers for the applications described in this manuscript. Plexin D1 antibody (Santa Cruz Biotechnology) used for magnetic force application experiments was validated for this application in-house.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

Bovine aortic endothelial cells were purchased from VEC Technologies; mouse endothelial cells used in this study were isolated in-house.

Authentication

Authentication was performed by western blotting and immunofluorescence for endothelial markers (PECAM, eNOS), as well as alignment in response to laminar flow.

Mycoplasma contamination

Cell lines used in this study were mycoplasma free.

Commonly misidentified lines (See [ICLAC](#) register)

None

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

All mice had a C57BL6/J background. Plexin D1fl/fl mice were obtained from J Epstein, Cdh5-Cre mice were obtained from R. Adams, ApoE^{-/-} mice were first generated and characterized by N. Maeda

Wild animals

No wild animals were used in this study.

Field-collected samples

No field-collected samples were used in this study.

Stress- and ubiquitylation-dependent phase separation of the proteasome

<https://doi.org/10.1038/s41586-020-1982-9>

Received: 23 September 2018

Accepted: 9 December 2019

Published online: 5 February 2020

Sayaka Yasuda^{1,8}, Hikaru Tsuchiya^{1,8}, Ai Kaiho^{1,8}, Qiang Guo², Ken Ikeuchi^{3,4}, Akinori Endo¹, Naoko Arai¹, Fumiaki Ohtake¹, Shigeo Murata⁵, Toshifumi Inada³, Wolfgang Baumeister², Rubén Fernández-Busnadiego^{2,6,7}, Keiji Tanaka^{1*} & Yasushi Saeki^{1*}

The proteasome is a major proteolytic machine that regulates cellular proteostasis through selective degradation of ubiquitylated proteins^{1,2}. A number of ubiquitin-related molecules have recently been found to be involved in the regulation of biomolecular condensates or membraneless organelles, which arise by liquid–liquid phase separation of specific biomolecules, including stress granules, nuclear speckles and autophagosomes^{3–8}, but it remains unclear whether the proteasome also participates in such regulation. Here we reveal that proteasome-containing nuclear foci form under acute hyperosmotic stress. These foci are transient structures that contain ubiquitylated proteins, p97 (also known as valosin-containing protein (VCP)) and multiple proteasome-interacting proteins, which collectively constitute a proteolytic centre. The major substrates for degradation by these foci were ribosomal proteins that failed to properly assemble. Notably, the proteasome foci exhibited properties of liquid droplets. RAD23B, a substrate-shuttling factor for the proteasome, and ubiquitylated proteins were necessary for formation of proteasome foci. In mechanistic terms, a liquid–liquid phase separation was triggered by multivalent interactions of two ubiquitin-associated domains of RAD23B and ubiquitin chains consisting of four or more ubiquitin molecules. Collectively, our results suggest that ubiquitin-chain-dependent phase separation induces the formation of a nuclear proteolytic compartment that promotes proteasomal degradation.

To enable visualization of proteasomes in live cells, we generated derivatives of the HCT116 colon cancer cell line in which endogenous proteasome subunits, the core particle subunit proteasome subunit β type-2 (PSMB2, also known as β 4), and regulatory particle subunit 26S proteasome non-ATPase regulatory subunit 6 (PSMD6, also known as RPN7), were labelled with an eGFP or FusionRed fluorescent tag (Extended Data Fig. 1). Consistent with previous studies^{9,10}, the proteasomes were primarily observed in the nucleoplasm and cytoplasm of highly proliferating cells. Over the course of a series of experiments, we unexpectedly observed that under hyperosmotic stress, proteasomes rapidly formed multiple foci in the nucleus (Fig. 1a, Supplementary Video 1). We confirmed that foci formed in wild-type (WT) HCT116 cells using an endogenous antibody against the proteasome (Fig. 1a, Extended Data Fig. 2a). Various osmolytes, sucrose, glucose and NaCl stimulated foci formation, and the osmolarity required to induce the response was 100 mOsmol l⁻¹, close to the value observed during physiological changes associated with type II diabetes¹¹ (Extended Data Fig. 2b). PSMD6–eGFP also formed foci, and the foci were stained with an activity-based probe, suggesting the involvement of active 26S proteasomes (Fig. 1a, Extended Data Fig. 2c). Consistently, a snapshot

obtained by cryo-electron tomography revealed the clustering of 26S proteasomes in the nucleus upon osmotic stimulation (Fig. 1b, Extended Data Fig. 2d). We also observed these foci in immortalized retinal pigment epithelial (RPE-1) cells and mouse embryonic stem (ES) cells, suggesting that hyperosmotic-stress-induced formation of proteasome foci is a universal phenomenon (Extended Data Fig. 2e).

Proteasome foci are sites of proteolysis

Several nuclear bodies are related to the proteasome, such as promyelocytic leukemia protein (PML) nuclear bodies and Cajal bodies^{12–14}, but these bodies did not colocalize with proteasome foci (Extended Data Fig. 3a, b). Instead, proteasome foci colocalized almost completely with lysine 48 (K48)-linked ubiquitin chains, a major proteolytic signal for the proteasome, but not with non-proteolytic K63-linked chains (Fig. 1c, Extended Data Fig. 3c). To obtain further functional insights, we performed time-lapse imaging of PSMB2–eGFP cells (Fig. 1d). Five minutes after the addition of 0.2 M sucrose, the number of foci increased to a maximum of around 30 per nucleus, and the diameter of the foci increased to approximately 500 nm within 30 min. Subsequently, the foci gradually disappeared over

¹Laboratory of Protein Metabolism, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan. ²Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Martinsried, Germany. ³Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan. ⁴Gene Center and Center for Integrated Protein Science Munich, Department of Biochemistry, University of Munich, Munich, Germany. ⁵Laboratory of Protein Metabolism, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan. ⁶Institute of Neuropathology, University Medical Center Göttingen, Göttingen, Germany. ⁷Cluster of Excellence 'Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells' (MBExC), University of Göttingen, Göttingen, Germany. ⁸These authors contributed equally: Sayaka Yasuda, Hikaru Tsuchiya, Ai Kaiho. *e-mail: tanaka-kj@igakuken.or.jp; saeki-ys@igakuken.or.jp

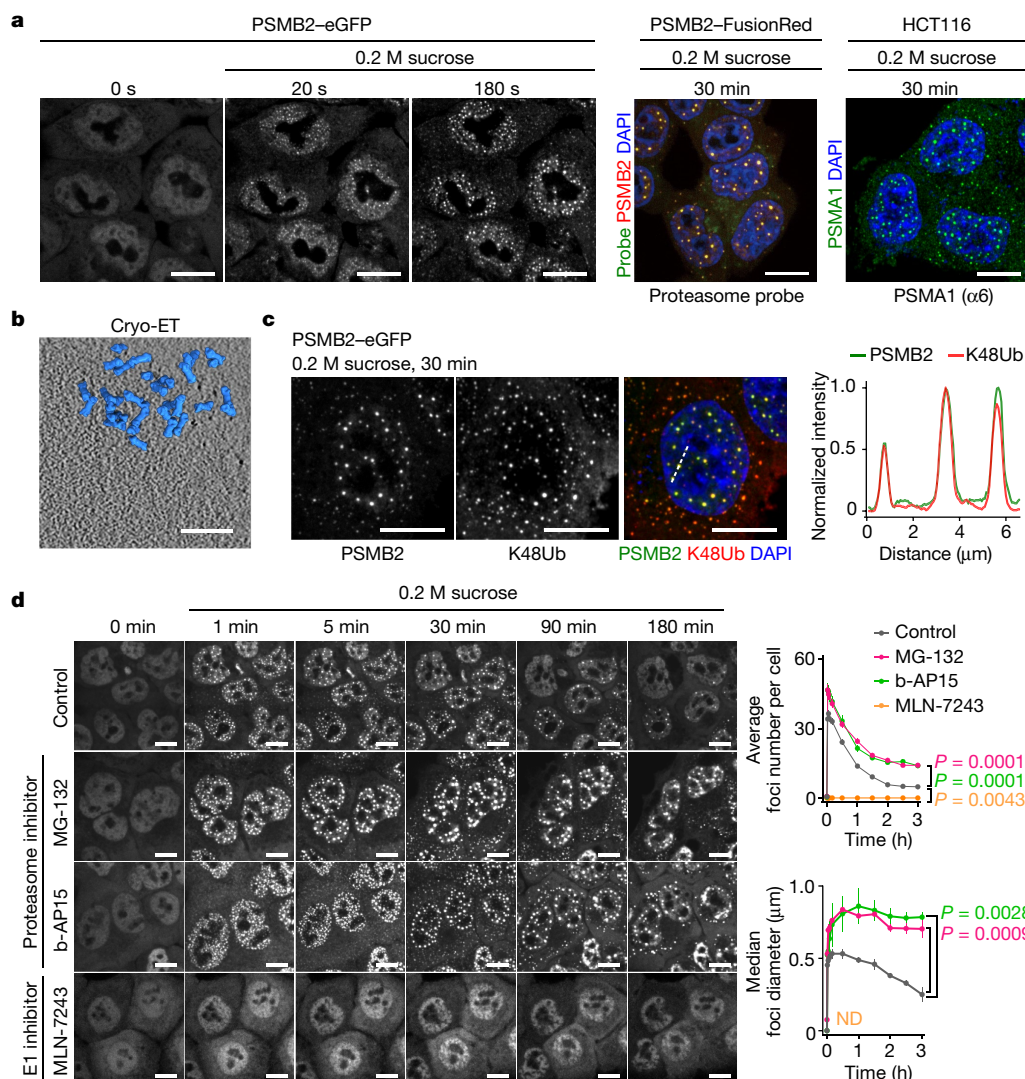


Fig. 1 | Hyperosmotic stress induces the formation of proteasome foci in the nucleus. **a**, HCT116 cells, PSMB2-eGFP or PSMB2-FusionRed cells were stimulated with 0.2 M sucrose. Endogenous proteasome activity was detected using a proteasome probe (Me4BodipyFL-Ahx3Leu3VS) and localization of endogenous proteasomes was determined using PSMA1 ($\alpha 6$) antibody at the indicated times. Scale bars, 10 μ m. **b**, Cryo-electron tomography (cryo-ET) image of proteasome foci in the nuclear region following stimulation with 0.2 M sucrose. Scale bar, 0.1 μ m. **c**, Left, PSMB2-eGFP cells were stimulated with 0.2 M sucrose for 30 min and endogenous K48Ub was detected with K48-ubiquitin antibody. Right, line profiling of a representative section of the cell, indicated by a white dashed line. Scale bars, 10 μ m. **d**, Left, time-lapse images of live PSMB2-eGFP cells. Cells were treated with proteasome inhibitor MG-132 (50 μ M, 1 h prior), b-AP15 (1 μ M, simultaneously) or ubiquitin-activating

enzyme (E1) inhibitor MLN-7243 (1 μ M, 1 h prior), and then stimulated with 0.2 M sucrose. Scale bars, 10 μ m. Representative images from two independent experiments. Right, changes of the proteasome foci number per cell and foci diameter in untreated control cells (grey) and MG-132- (magenta), b-AP15- (green) or MLN-7243- (orange) treated cells. n represents cell numbers (control: $n = 119, 121, 116, 117, 120, 120, 123, 126, 128, 125, 125$; MG132: $n = 94, 99, 99, 95, 95, 91, 93, 92, 94, 100$ and 95 ; b-AP15: $n = 81, 85, 80, 80, 85, 88, 90, 93, 103, 93$ and 102 ; MLN-7243: $n = 47, 43, 43, 42, 42, 42, 42, 40, 39, 39$ and 39 , at 0, 1, 3, 5, 10, 30, 60, 90, 120, 150 and 180 min after sucrose treatment, respectively). Data are mean \pm s.e.m.; one-way ANOVA with Dunnett's test (number) and median \pm s.e.m., Friedman with Dunn's test (diameter). P values are shown in the figure. Representative images from four (**a**, left), three (**a**, middle and right) or two (**b**, **d**) independent experiments.

the course of 3 h. When the cells were treated with the proteasome inhibitor MG-132 or b-AP15, the number and size of the foci increased, and clearance was significantly delayed. By contrast, pre-treatment with the ubiquitin E1 inhibitor MLN-7243 almost completely inhibited foci formation (Fig. 1d, bottom, Extended Data Fig. 3d). Thus, ubiquitylated substrates are required for the formation of proteasome foci and proteasome activity is necessary for their clearance. Together, these observations suggested that hyperosmotic-stress-induced proteasome foci are sites of proteolysis.

Orphan RP is degraded in proteasome foci

A slight increase in the level of ubiquitylated proteins was observed following hyperosmotic stimulation, and a mass spectrometry-based

ubiquitylome analysis identified changes in several housekeeping proteins, including linker histones, HSP90 and ribosomal proteins (RPs) (Extended Data Fig. 4a, b). We subsequently focused on RPs. Ribosomes are constitutively produced at around 7,500 molecules per minute in cultured human cells via complicated assembly processes within the nucleus, and orphan RPs that fail to incorporate into ribosomes are degraded by the ubiquitin-proteasome system^{15–19}. Therefore, we hypothesized that ribosome biosynthesis might be vulnerable to hyperosmotic stress. Indeed, electron microscopy analysis showed the disappearance of the nucleolar dense fibrillar compartment (DFC), where pre-ribosomal RNA (rRNA) and RPs are assembled, upon hyperosmotic stress (Fig. 2a). We also observed the emergence of multiple hyperdense structures near proteasome foci; these structures disappeared

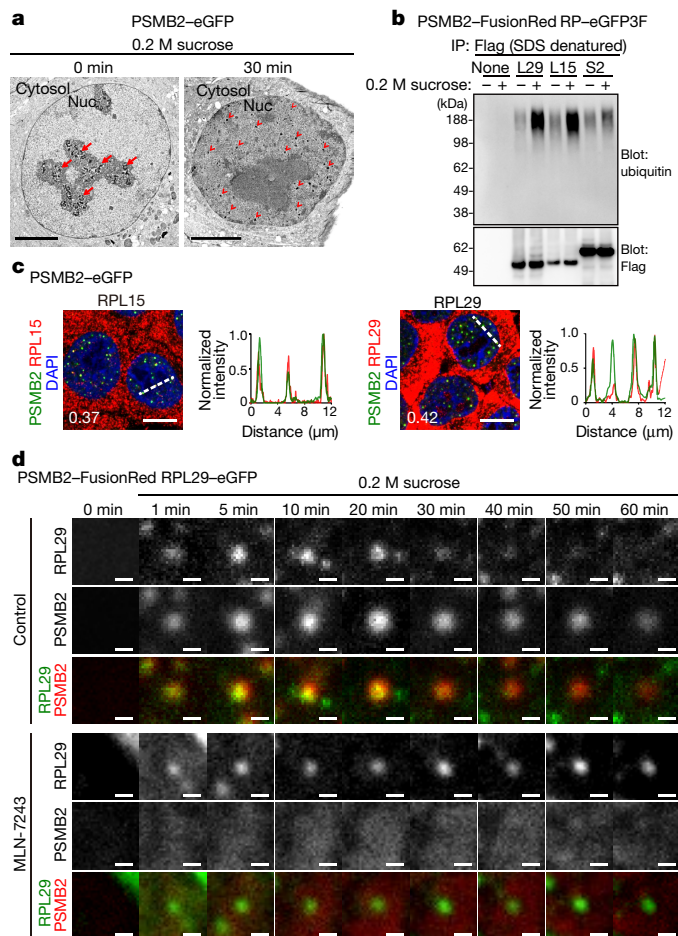


Fig. 2 | Ribosomal proteins are degraded in proteasome foci. **a**, Changes in the ultrastructural appearance of type 2 granule matrix in the nucleolus and nucleoplasm, as determined by transmission electron microscopy. Dense fibrillar compartment structures (arrow) in the nucleolus disappeared and very dense granules (arrowhead) in the nucleoplasm were observed after stimulation of PSMB2-eGFP cells with 0.2 M sucrose. Nuc, nucleus. Scale bars, 5 μ m. Representative images from two independent experiments. **b**, PSMB2-FusionRed cells stably expressing eGFP-3 \times Flag-fused RPL29 (L29), RPL15 (L15) or RPS2 (S2) were stimulated with or without 0.2 M sucrose for 30 min, immunoprecipitated (IP) with Flag antibody and immunoblotted as indicated. Representative result from three independent experiments. For gel source data, see Supplementary Fig. 1. **c**, Colocalization of proteasome foci and endogenous ribosomal proteins. PSMB2-eGFP cells were stimulated with 0.2 M sucrose for 30 min, and endogenous ribosomal proteins were detected with specific antibodies. The mean value of the Pearson correlation coefficient in the nucleoplasm is shown in the image ($n = 10$ cells in two fields of view). Scale bars, 10 μ m. Each graph represents the normalized fluorescence distribution over the white dashed lines. Representative results from two independent experiments. See Extended Data Fig. 5a for additional ribosomal components. **d**, Time-lapse images of single foci in live HCT116 cells stably expressing PSMB2-FusionRed and eGFP-RPL29 (PSMB2-FusionRed/RPL29-eGFP). Cells were stimulated with 0.2 M sucrose, with or without pretreatment with the E1 inhibitor MLN-7243 (1 μ M, 1 h prior). Scale bars, 0.5 μ m. Representative results from five (control) or three (MLN-7243) independent experiments. See also Supplementary Videos 2 and 3.

in a proteasome activity-dependent manner after 4 h (Extended Data Fig. 4c, d). Northern blot analysis revealed that hypertonic stress caused a reduction in the Pol I-transcribed pre-rRNA and its processed forms¹⁵ (Extended Data Fig. 4e). In addition, stably expressed RPs in cells were constitutively ubiquitinated, and notably, ubiquitylation levels were further increased by sucrose treatment (Fig. 2b). Thus, hyperosmotic

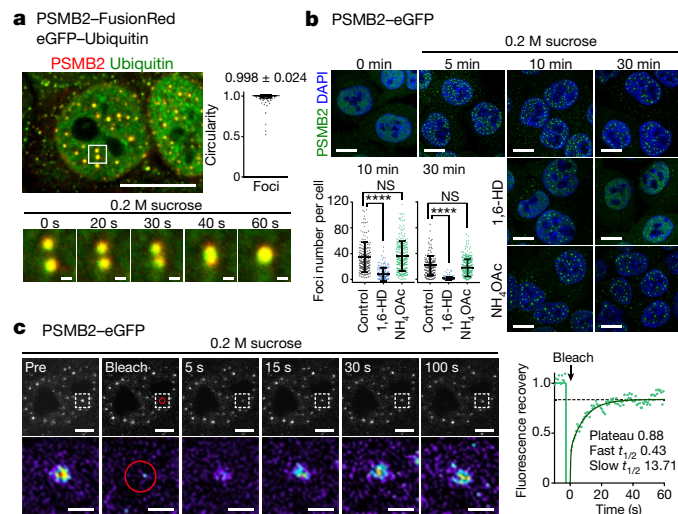


Fig. 3 | Proteasome foci are liquid droplets. **a**, Foci fusion in living HCT116 cells stably expressing PSMB2-FusionRed and eGFP-ubiquitin after stimulation with 0.2 M sucrose. Scale bar, 10 μ m. Bottom, enlarged time-course views of the square in the above image. Scale bars, 0.5 μ m. The graph indicates the circularity of individual foci 30 min after sucrose stimulation. Data are mean \pm s.d., $n = 1,312$ foci from 98 cells in two fields of view. **b**, PSMB2-eGFP cells were treated with 1% 1,6-hexanediol (1,6-HD) or 0.1 M NH_4OAc 2 min after induction of foci formation by 0.2 M sucrose. Scale bars, 10 μ m. The graph shows the number of foci per cell at the indicated times. n represents cell numbers (control: 166 cells (10 min), 147 cells (30 min); 1,6-HD, 147 cells (10 min), 174 cells (30 min); NH_4OAc , 223 cells (10 min), 283 cells (30 min)). Data are mean \pm s.d. and were analysed by Kruskal–Wallis with Dunn’s multiple comparison test. $P = 0.7949$ (NH_4OAc , 10 min), $P = 0.2274$ (NH_4OAc , 30 min), **** $P < 0.0001$. NS, not significant. **c**, Left, FRAP (red outline, 1 μ m circle) in a 0.2 M sucrose-induced proteasome focus in a PSMB2-eGFP cell. Scale bars, 5 μ m (top) and 1 μ m (bottom). Right, quantification of fluorescence recovery of the proteasome focus and double-exponential fitting curve (dark green). Representative results from six independent experiments.

stress induces acute nucleolar stress, thereby causing failure of pre-ribosome assembly.

To determine whether RPs localize to proteasome foci, we performed immunofluorescence staining, and found that several RPs (RPL7A, RPL15, RPL29 and RPS2) modestly colocalized with proteasome foci (Fig. 2c, Extended Data Fig. 5a). In time-lapse imaging of the PSMB2-FusionRed cells stably expressing RPL29-eGFP, we observed a rapid emergence of RPL29 condensates and their degradation at proteasome foci (Fig. 2d, Extended Data Fig. 5b, Supplementary Video 2). RPL29 structures were also observed in the presence of MLN-7243, but these were smaller, and their clearance was markedly delayed (Fig. 2d, Supplementary Video 3). Treatment with the Pol I inhibitor CX-5461 or transient overexpression of RPs further increased the size of proteasome foci, suggesting that the condensates formed from unassembled orphan RPs produced in the nucleoplasm (Extended Data Fig. 5c, d).

Proteasome foci are liquid droplets

Time-lapse imaging of the PSMB2-FusionRed cells stably expressing eGFP-ubiquitin revealed that proteasome foci have liquid droplet-like properties: small foci that contained proteasomes and ubiquitin suspended in the nucleoplasm fused into larger foci, and their circularity was 0.998 (Fig. 3a, Supplementary Video 4). The number of preformed foci decreased upon addition of 1,6-hexanediol, an aliphatic alcohol that destabilizes liquid droplets, but not by ammonium acetate, which disrupts RNA gelation^{20,21} (Fig. 3b). Fluorescence recovery after photobleaching (FRAP) experiments revealed that approximately 90% of

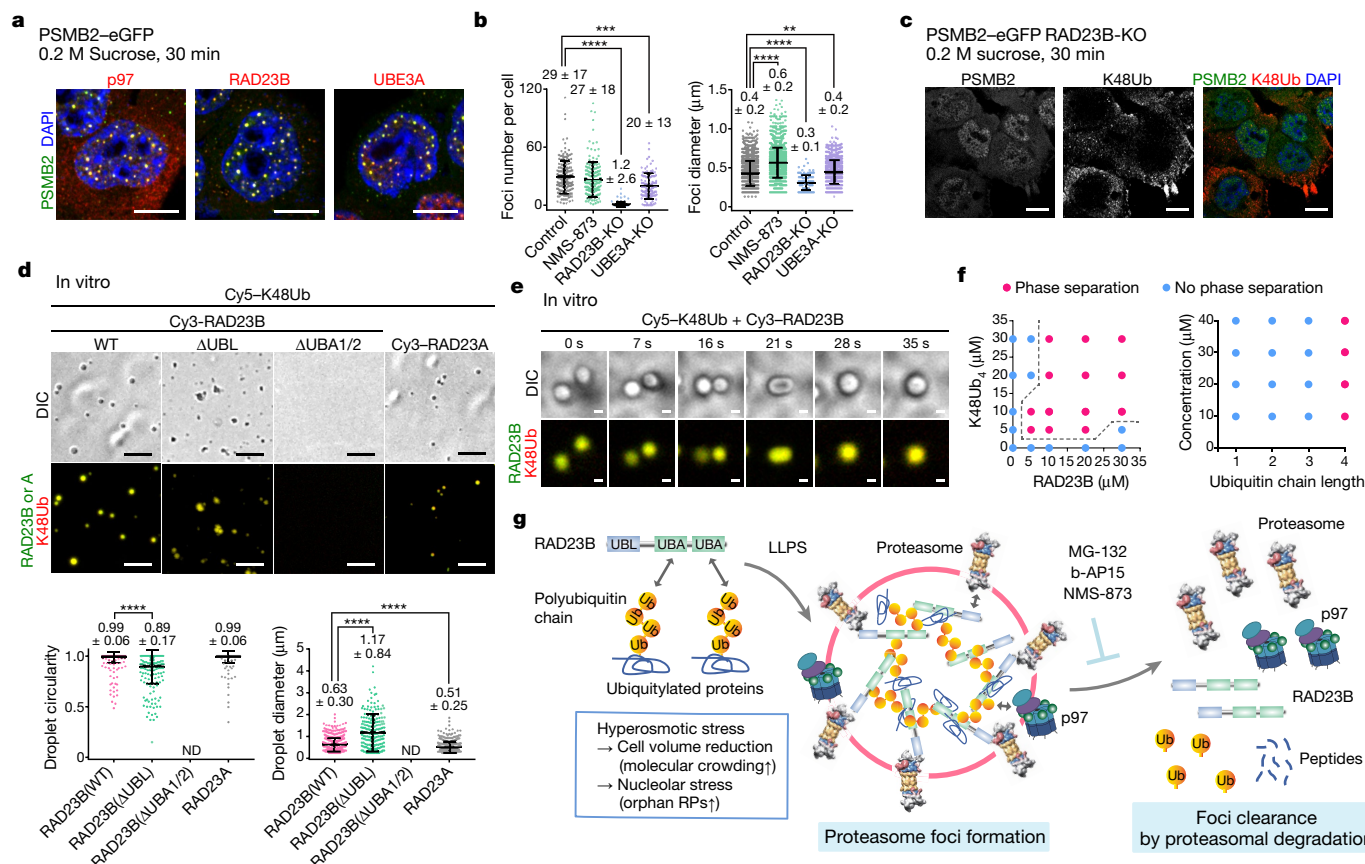


Fig. 4 | Polyubiquitin chain and RAD23B induce liquid-liquid phase separation of the proteasome. **a**, p97, RAD23B and UBE3A localized to proteasome foci. PSMB2-eGFP cells were stimulated with 0.2 M sucrose and observed by immunofluorescence with indicated antibodies. Scale bars, 10 μm. **b**, PSMB2-eGFP cells treated with p97 inhibitor NMS-873 (1 μM, 1 h prior), PSMB2-eGFP cells lacking RAD23B or UBE3A were stimulated with 0.2 M sucrose for 30 min. The graph indicates the number of proteasome foci per cell and the diameter of individual foci. *n* represents cell numbers (control, 176 cells; NMS-873, 168 cells; RAD23B-KO, 112 cells; UBE3A-KO, 108 cells). Data are mean ± s.d., *****P* < 0.0001, ****P* < 0.0002, ***P* < 0.0021 by Kruskal-Wallis with Dunn's test. **c**, RAD23B-KO PSMB2-eGFP (PSMB2-eGFP/RAD23B-KO) cells were stimulated with 0.2 M sucrose for 30 min, and endogenous K48Ub was detected with K48-ubiquitin antibody. Scale bars, 10 μm. **d**, Top, liquid droplets formed 90 min after mixing of 20 μM Cy5-K48Ub chains with 20 μM Cy3-

RAD23B(WT), Cy3-RAD23B UBL deletion mutant (ΔUBL), Cy3-RAD23B UBA1 and UBA2 deletion mutant (ΔUBA1/2) or Cy3-RAD23A in 3% PEG and 200 mM sodium chloride. Scale bars, 5 μm. Bottom, quantified circularity and quantified diameter of individual droplets (RAD23B(WT), *n* = 691 droplets; ΔUBL, *n* = 230 droplets; RAD23A, *n* = 491 droplets). Data are mean ± s.d., *****P* < 0.0001 by Kruskal-Wallis with Dunn's test. **e**, Fusion of liquid droplets containing Cy3-RAD23B and Cy5-K48Ub in vitro. The sample was identical to the one used in **d** (15 min after mixing in 10% PEG). Scale bars, 1 μm. **f**, Left, phase diagram of RAD23B and K48-tetraubiquitin (K48Ub₄) at different concentrations. Right, phase diagram of RAD23B (10 μM) and ubiquitin chains with different lengths. Red dots indicate phase separation and blue dots indicate no phase separation. **g**, Current model of formation and clearance of proteasome-containing liquid droplets. In **a**, **c**–**e**, representative results from two independent experiments.

proteasomes were rapidly exchanged into and out of the foci, further supporting the idea that proteasome foci are liquid droplets (Fig. 3c).

p97 and RAD23B regulate proteasome foci

A mass-spectrometry-based proteomic screen for proteasome-interacting proteins identified ubiquitin-selective chaperone p97 (also known as VCP), substrate-shuttling factor RAD23B and ubiquitin ligase UBE3A (also known as E6-AP)¹ (Source Data of Fig. 4a). These proteins extensively colocalized with the proteasome at the foci (Fig. 4a, Extended Data Fig. 6a, b). Inhibition of p97 activity by NMS-873 caused an increase of around 50% in foci size, suggesting a positive role of p97 for degradation of ubiquitylated substrates in proteasome foci (Fig. 4b). Indeed, NMS-873, like b-AP15, increased the size of RPL29 condensates (Extended Data Fig. 7, Supplementary Videos 5, 6). Knockout (KO) of UBE3A caused a reduction of around 30% in the number of foci, suggesting that UBE3A regulates the proteasome itself or ubiquitylation of substrate proteins in the foci²² (Fig. 4b, Extended Data Fig. 8a). Notably, formation of proteasome foci was markedly

attenuated in RAD23B-KO cells. No significant effects were observed on siRNA-mediated knockdown of ubiquitin-like proteins (UBQLNs) and RAD23A, other shuttling factors, or XPC, which functions with RAD23 proteins in the DNA nucleotide excision repair pathway^{1,23} (Extended Data Fig. 6c).

We initially predicted that ubiquitylated proteins would form foci before the RAD23B-dependent recruitment of proteasomes. However, we found that neither the proteasome nor ubiquitin formed the foci in RAD23B-KO cells (Fig. 4c, Extended Data Fig. 8b). This was not due to a decrease in the level of ubiquitylated proteins (Extended Data Fig. 8a). RAD23 family proteins have a ubiquitin-like (UBL) domain and two ubiquitin-associated (UBA) domains that bind the proteasome and ubiquitin chains, respectively^{1,24}. Adding back mutant RAD23B to RAD23B-KO cells revealed that formation of ubiquitin-positive foci was UBA domain-dependent (Extended Data Fig. 8b). Because overexpression of RAD23A rescued RAD23B-KO cells, the difference in ability to form foci can be explained by differences in the expression levels of the two proteins^{25,26} (Extended Data Fig. 8c, d).

RAD23B drives LLPS of ubiquitin chains

These results raised the possibility that RAD23B is directly involved in liquid–liquid phase separation (LLPS) of ubiquitylated proteins. To determine this *in vitro*, we incubated fluorescently labelled RAD23B and K48-linked polyubiquitin (K48Ub) chains in the presence of polyethylene glycol (PEG), used as a crowding reagent (Extended Data Fig. 9a–c). On mixing, RAD23B and K48Ub chains formed spherical condensates in which the two different fluorescent signals were uniformly distributed (Fig. 4d). FRAP analysis revealed rapid exchanges of both RAD23B and K48Ub chains, suggesting co-phase separation of RAD23B and K48Ub chains (Extended Data Fig. 9d). Consistent with the properties of liquid droplets, small condensates fused into larger ones of up to micrometre size (Fig. 4e, Supplementary Video 7). RAD23B lacking the UBL domain caused formation of amorphous protein aggregates, whereas RAD23B lacking the UBA domains did not form condensates at all (Fig. 4d). As RAD23B prefers K48Ub chains with four or more ubiquitin molecules^{25,27}, co-phase separation of RAD23B/K48Ub chains was dependent not only on the concentration of each protein but also on the length of K48Ub chains (Fig. 4f, Extended Data Fig. 9e). Although long K63-linked ubiquitin chains could form condensates with RAD23B, the efficiency was lower than that of K48Ub chains (Extended Data Fig. 9e). Thus, multivalent interactions between long K48Ub chains and two UBA domains of RAD23B drive liquid–liquid phase separation.

Discussion

In this study, we identified a proteasome-containing structure that is induced by hyperosmotic stress. The fluid organization arises from LLPS of ubiquitylated proteins and RAD23B, followed by proteasome recruitment. Proteasome condensates were prominently observed in the nucleoplasm, probably because hyperosmotic stress results in a further increase in the nuclear concentration of proteasomes, RAD23B and ubiquitylated substrates, and in particular, ubiquitylated orphan RPs, owing to nucleolar stress (Fig. 4g). Although its functional importance is not fully understood, the condensation appears to facilitate proteasomal degradation, because ribosomal condensates were stabilized by inhibition of the proteasome or p97 (Extended Data Fig. 7). Moreover, in RAD23B-KO cells, as in cells treated with E1 inhibitor, small amorphous structures of RPL29 were observed, suggesting that condensation of ubiquitylated proteins might protect against protein aggregation (Extended Data Fig. 7). Given that unassembled RPs stimulate p53 activation²⁸, failure of ribosomal condensate formation might cause apoptosis. Indeed, RAD23B-KO cells underwent apoptosis in response to mild hyperosmotic stress (Extended Data Fig. 8e, f). Conversely, recent studies showed a conversion from liquid-like droplets to solid-like assemblies of aggregation prone proteins, most of which the proteasome can degrade only in reversible aggregated forms^{3,4,29,30}. In this context, acute hyperosmotic stress may risk irreversible accumulation of protein aggregates, especially when the proteasome or p97 activity is reduced.

It remains unclear whether multivalent interactions between ubiquitin chains and ubiquitin-binding proteins universally induce LLPS in cells. Given that cells contain numerous ubiquitin-binding proteins that regulate multiple cellular pathways, and in light of the profound functional consequences of biomolecular condensation, it will be of great interest to investigate their ability to promote LLPS of ubiquitylated proteins as well as their physiological consequences.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1982-9>.

1. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
2. Livneh, I., Cohen-Kaplan, V., Cohen-Rosenzweig, C., Avni, N. & Ciechanover, A. The life cycle of the 26S proteasome: from birth, through regulation and function, and onto its death. *Cell Res.* **26**, 869–885 (2016).
3. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
4. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
5. Bouchard, J. J. et al. Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, phase-separated compartments. *Mol. Cell* **72**, 19–36 (2018).
6. Turakhiya, A. et al. ZFAND1 recruits p97 and the 26S proteasome to promote the clearance of arsenite-induced stress granules. *Mol. Cell* **70**, 906–919 (2018).
7. Dao, T. P. et al. Ubiquitin modulates liquid-liquid phase separation of UBQLN2 via disruption of multivalent interactions. *Mol. Cell* **69**, 965–978 (2018).
8. Sun, D., Wu, R., Zheng, J., Li, P. & Yu, L. Polyubiquitin chain-induced p62 phase separation drives autophagic cargo segregation. *Cell Res.* **28**, 405–415 (2018).
9. Wójcik, C. & DeMartino, G. N. Intracellular localization of proteasomes. *Int. J. Biochem. Cell Biol.* **35**, 579–589 (2003).
10. Enekel, C. Proteasome dynamics. *Biochim. Biophys. Acta* **1843**, 39–46 (2014).
11. Umpierrez, G. & Korytkowski, M. Diabetic emergencies—ketoacidosis, hyperglycaemic hyperosmolar state and hypoglycaemia. *Nat. Rev. Endocrinol.* **12**, 222–232 (2016).
12. Janer, A. et al. PML clastosomes prevent nuclear accumulation of mutant ataxin-7 and other polyglutamine proteins. *J. Cell Biol.* **174**, 65–76 (2006).
13. Cioce, M., Boulon, S., Matera, A. G. & Lamond, A. I. UV-induced fragmentation of Cajal bodies. *J. Cell Biol.* **175**, 401–413 (2006).
14. Levy-Barda, A. et al. Involvement of the nuclear proteasome activator PA28 γ in the cellular response to DNA double-strand breaks. *Cell Cycle* **10**, 4300–4310 (2011).
15. Bohnsack, K. E. & Bohnsack, M. T. Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J.* **38**, e100278 (2019).
16. Lam, Y. W., Lamond, A. I., Mann, M. & Andersen, J. S. Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr. Biol.* **17**, 749–760 (2007).
17. Sung, M. K. et al. A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins. *eLife* **5**, e19105 (2016).
18. Nguyen, A. T. et al. UBE2O remodels the proteome during terminal erythroid differentiation. *Science* **357**, eaan0218 (2017).
19. Yanagitani, K., Juszkiwicz, S. & Hegde, R. S. UBE2O is a quality control factor for orphans of multiprotein complexes. *Science* **357**, 472–475 (2017).
20. Kroschwald, S. et al. Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *eLife* **4**, e06807 (2015).
21. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
22. Jacobson, A. D., MacFadden, A., Wu, Z., Peng, J. & Liu, C. W. Autoregulation of the 26S proteasome by *in situ* ubiquitination. *Mol. Biol. Cell* **25**, 1824–1835 (2014).
23. Yokoi, M. & Hanaoka, F. Two mammalian homologs of yeast Rad23, HR23A and HR23B, as multifunctional proteins. *Gene* **597**, 1–9 (2017).
24. Walters, K. J., Lech, P. J., Goh, A. M., Wang, Q. & Howley, P. M. DNA-repair protein hHR23a alters its protein structure upon binding proteasomal subunit S5a. *Proc. Natl Acad. Sci. USA* **100**, 12694–12699 (2003).
25. Nathan, J. A., Kim, H. T., Ting, L., Gygi, S. P. & Goldberg, A. L. Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? *EMBO J.* **32**, 552–565 (2013).
26. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
27. Kristariyanto, Y. A. et al. K29-selective ubiquitin binding domain reveals structural basis of specificity and heterotypic nature of K29 polyubiquitin. *Mol. Cell* **58**, 83–94 (2015).
28. Liu, Y., Deisenroth, C. & Zhang, Y. RP-MDM2–p53 pathway: linking ribosomal biogenesis and tumor surveillance. *Trends Cancer* **2**, 191–204 (2016).
29. Dikic, I. Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).
30. Klaijs, C. L., Jayaraj, G. G. & Hartl, F. U. Pathways of cellular proteostasis in aging and disease. *J. Cell Biol.* **217**, 51–63 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture

HCT116 cells (human colorectal carcinoma, CCL-247) and hTERT RPE-1 cells (human telomerase-immortalized retinal pigmented epithelial, CRL-4000), both obtained from ATCC (Manassas) were maintained at 37 °C in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% (v/v) fetal bovine serum (FBS) (Biowest), 1 mM sodium pyruvate (Thermo Fisher Scientific), and nonessential amino acids (Thermo Fisher Scientific) in an incubator with a 5% CO₂ atmosphere. E14TG2a mouse ES cells purchased from ATCC (CRL-1821) were cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% (v/v) Embryonic Stem-Cell FBS (Thermo Fisher Scientific), 1 mM sodium pyruvate, nonessential amino acids, 0.1 mM 2-ME (Sigma), 15% (v/v) FBS-free medium supplement (Thermo Fisher Scientific), and 10³ U ml⁻¹ unit ESGRO Leukaemia Inhibitory Factor (Merck).

Inhibitors and reagents

Formation of proteasome foci was induced by NaCl (Wako), sucrose (Wako) or glucose (Wako) at the indicated concentrations. Me4BodipyFL-Ahx3Leu3VS fluorescent proteasome probe³¹ (UbiQ Bio) was used to detect endogenous proteasome activity. PSMB2-FusionRed^{KI/KI} cells were treated with 1 μM Me4BodipyFL-Ahx3Leu3VS for 1 h and washed with PBS, after which formation of proteasome foci was induced with 0.2 M sucrose. Ammonium acetate (Wako) and 1,6-hexanediol (Sigma) were used to distinguish liquid-like and solid-like states of proteasome foci in living cells. Hexanediol (1%) and 0.1 M ammonium acetate were added 2 min after induction of foci formation by 0.2 M sucrose in PSMB2-eGFP^{KI/KI} cells. Because hexanediol can cause various deleterious effects on cells, treatment with this reagent was minimized³². The inhibitors used in this study were as follows: bortezomib (LC Laboratories), MG-132 (Peptide Institute), b-AP15³³ (LifeSensors), MLN-7243³⁴ (Active Biochem), NMS-873³⁵ (Selleck Chemicals), DBeQ (Sigma) and CX-5461 (ChemScene).

MTS assay for cell viability

Cells treated with different concentration of DBeQ for 4 h, and cell viability was analysed by CellTiter 96 AQueous One Solution Cell Proliferation Assay kit (Promega). Absorbance at 490 nm was measured on an EnSpire 2300 multimode plate reader (PerkinElmer).

Plasmids and plasmid transfection

Transfections of HCT116 cells with the indicated plasmids were performed using Lipofectamine 2000 (Thermo Fisher Scientific). Human RPL7A and RPS2 were cloned into pcDNA3-eGFP (Addgene 13031) with a GGS linker sequence and 3×Flag to yield pcDNA3-RPL7A-eGFP-3×Flag and pcDNA3-RPS2-eGFP-3×Flag, respectively. To generate pcDNA3-RAD23A-FusionRed-3×Flag, human RAD23A was cloned with a GGS linker sequence and 3×Flag into pcDNA3-eGFP in which the eGFP gene was replaced with FusionRed (Evrogen, JSC). To construct donor vectors for TALEN knock-in (KI), -1,000 bp upstream of the target site, including the last exon without the stop codon, a linker sequence (GGS linker), eGFP or FusionRed, 3×Flag tag, SV40 polyA, selection marker, and -1,000 bp of downstream sequence were tandemly flanked by PCR and inserted into pBlueScript (Stratagene).

RNA interference

siRNA oligos were obtained from ON-TARGETplus SMARTpool (Dharmacon). The target sequences are as follows: RAD23A#1, GCTCTGAGTATGAGACGAT; RAD23A#2, GAAGATAGAAGCTGAGAAG; RAD23A#3, GATCTTGAGTGACGATGTC; RAD23A#4, GAAGAACTTTGTGGTCGTC;

RAD23B#1, GCAGATAGGTCGAGAGAAT; RAD23B#2, GAACGAGAGCAAGTAATTG; RAD23B#3, GAAAGTGGTCATATGAACCTA; RAD23B#4, CAACAACCCTGACAGAGCA; UBQLN1#1, CATCAACTCCTAATAGTAA; UBQLN1#2, GTACTACTGCGCCAAATTT; UBQLN1#3, AGACAAACGTTGGAACCTG; UBQLN1#4, CTGAGTAGCTTGGGTTTGA; UBQLN2#1, TAAGGAAGCGATTTCGAAA; UBQLN2#2, CTGAATAGCCCGCTGTTTA; UBQLN2#3, CCAAACCGATCAGCTAGTG; UBQLN2#4, GACATTAGCCACTGAAGCA; UBQLN4#1, GCTGAGAATATGACGGCAA; UBQLN4#2, CCAATGAAGCTAAGCGCCA; UBQLN4#3, GAGGAGGGAATGAGATTAT; UBQLN4#4, CCAACCAATGCTAGAATTT; PML#1, GGAAAGATGCAGCTGTATC; PML#2, GAGCTCAAGTGCAGCATCA; PML#3, GGACATGCACGGTTTCCTG; PML#4, GCAACCAAGTGCAGTGCCTGA; XPC#1, GCAAATGGCTTCTATCGAA; XPC#2, TGAAATATGAGGCCATCTA; XPC#3, GAGAAGTACCCTACAAGAT; XPC#4, GGAGGGCGATGAAACGTTT. For non-targeting control siRNA, we used ON-TARGETplus non-targeted siRNA#4 (Dharmacon). siRNAs were transfected into cells using Lipofectamine RNAiMAX (Thermo Fisher Scientific). After 24 h of transfection, the medium was replaced, and the cells were grown for an additional 24 h before analysis.

Generation of KI cell lines

PSMB2-eGFP-3×Flag- or PSMB2-FusionRed-KI HCT116 cells were generated by TALEN KI using the Platinum Gate TALEN Kit, a gift from T. Yamamoto (Addgene# 1000000043)³⁶. TALEN-targeting sequences (PSMB2-1: TTTCTTCCCAAACAGGGCT; PSMB2-2: TTCCCTGGCAAGTGGGAGGGA) at the last exon of *PSMB2* were cloned into pCMV-153/47. The assembled TALEN plasmids and a donor vector pcDNA3-PSMB2-linker-eGFP-3×Flag (Puro) or pcDNA3-PSMB2-linker-FusionRed (Puro) were co-transfected into HCT116 cells. Puromycin-resistant clones were isolated and validated by western blot and DNA sequencing. PSMD6-eGFP-3×Flag-KI HCT116 cells were generated using the TALEN nuclease (TALEN) Kit³⁷, a gift from F. Zhang (Addgene 1000000019). TALEN-targeting sequences (PSMD6-1: TCCAGAGTAATTAATATGTA; PSMD6-2: TCCAGAGTAATTAATATGTA) at the last exon of *PSMD6* were cloned into pTALEN_v2. The assembled TALEN plasmids and the donor vector, pBlueScript-PSMD6-linker-eGFP-3×Flag (Puro), were co-transfected into HCT116 cells, and puromycin-resistant clones were isolated and validated. eGFP-ubiquitin, RPL29-eGFP, RPL15-eGFP or RPS2-eGFP was knocked in to the *AAVS1* locus of HCT116 cells using the TALEN Kit using TALEN-targeting sequences AAVS1-1 (TGCCCTCCACCCACA) and AAVS1-2 (TTTCTGTCCACCAATCCTG). The TALEN plasmids and a donor vector pBlueScript-EF1pr-EGFP-ubiquitin (Neo), pBlueScript-EF1pr-RPL15-linker-eGFP-3×Flag (Neo), pBlueScript-EF1pr-RPL29-linker-eGFP-3×Flag (Neo) or pBlueScript-EF1pr-RPS2-linker-eGFP-3×Flag (Neo) were co-transfected into HCT116 cells. G418-resistant clones were isolated and validated by immunoblot and DNA sequencing.

Generation of UBE3A-KO and RAD23B-KO cells

A CRISPR guide sequence targeting exon 3 of *UBE3A* or exon 5 of *RAD23B* was cloned into pSpCas9(BB)-2A-Puro (PX459) v.2.0³⁸, a gift from F. Zhang (Addgene 62988). Sequences were as follows: UBE3A-1: CACCGCTTACCTTGAGAACTCGAA; UBE3A-2: AAACCTTCGAGTTCTCAAGTAAGC; RAD23B-1: CACCGAAGATGCAACGAGTGCACT; RAD23B-2: AAACAGTGCATCGTTGCATCTTC. Puromycin-resistant clones were isolated and validated by western blotting and DNA sequencing.

Generation of RAD23B add-back cell lines

RAD23B add-back cell lines were generated using a retroviral system. For expression of human RAD23B, RAD23B mutants, L8A and L225A/L401A, were generated by PCR mutagenesis using the QuikChange mutagenesis kit (Agilent Technologies) and cloned into pMX-Puro retroviral expression vector (Cell Biolabs) to yield pMX-puro-RAD23B(WT)-FusionRed-3×Flag, pMX-puro-RAD23B(L8A)-FusionRed-3×Flag, and pMX-puro-RAD23B(L225A/L401A)-FusionRed-3×Flag, respectively.

Article

Virus particles were produced in HEK293T cells co-transfected with Gag-Pol, VSV-G, and RAD23B retrovirus plasmids in six-well plates. After 12 h of transfection, the medium was replaced and the cells were cultivated for an additional 24 h. Viral supernatants were then used to infect RAD23B-KO cells. Puromycin-resistant clones were isolated and validated by western blotting.

Microscopy

Live-cell imaging experiments were performed on a CV1000 automated spinning-disk microscope (Yokogawa Electric Corporation) equipped with a UPLSApo60 × O 1.35NA (Olympus), or on an IX73 inverted fluorescence microscope (Olympus) equipped with an enhanced CSU-X1 spinning disk (Microlens-enhanced dual Nipkow disk confocal scanner) (Yokogawa), a PlanApo 100 × OTIRF 1.45NA (Olympus), and an Andor Neo sCMOS camera (Andor) or an ORCA-Flash4.0 V3 Digital CMOS camera (Hamamatsu Photonics). All cells were maintained in a humidified environment at 37 °C under 5% CO₂. Three-dimensional imaging experiments were performed on a Leica TCS SP8 laser-scanning microscope (Leica Microsystems) with a HC PL APO 100×/1.40 NA oil CS2 (Leica). In vitro droplet formation of protein samples was monitored on a Fluoview FV3000 confocal laser scanning microscope with an IX83 fully motorized inverted microscope and a UAPON 100× OTIRF 1.49 NA (Olympus).

Immunofluorescence and time-lapse imaging

Cells were initially plated in 35-mm glass-bottomed dishes (MatTek) coated with poly-L-lysine. For immunofluorescence, cells were fixed in 4% paraformaldehyde (Thermo Fisher Scientific) in PBS for 15 min, and then permeabilized with MeOH (Wako) for 5 min at −20 °C. Before antibody incubation, cells were blocked with 1% FBS in PBS. The primary antibodies used were as follows: anti-multi-ubiquitin mouse monoclonal FK2 (NBT-MFK003; Nippon Bio-Test Laboratories), anti-multi-ubiquitin rabbit polyclonal (Z0458; Dako), Lys48-Specific anti-ubiquitin rabbit monoclonal (05-1307; clone Apu2; Millipore), or Lys63-Specific anti-ubiquitin rabbit monoclonal (05-1308; clone Apu3; Millipore); anti-PSMA1 mouse monoclonal antibody³⁹; anti-RAD23B rabbit monoclonal (13525; Cell Signaling Technology); anti-VCP mouse monoclonal (ab11433; Abcam); anti-UBE3A rabbit polyclonal (10344-1-AP; Protein Tech); anti-RPS2 rabbit polyclonal (ab155961; Abcam); anti-RPS6 rabbit monoclonal (2217; Cell Signaling Technology); anti-RPS9 rabbit polyclonal (18215-1-AP; Protein Tech); anti-RPL4 mouse monoclonal (sc-100838; Santa Cruz Biotechnology); anti-RPL7A rabbit polyclonal (2415; Cell Signaling Technology); anti-RPL15 rabbit polyclonal (16740-1-AP; Protein Tech); anti-RPL29 mouse polyclonal (H00006159-B01P; Abnova); anti-RPL35 rabbit polyclonal (SAB4500233; Sigma-Aldrich); anti-rRNA mouse polyclonal (sc-33678; Santa Cruz Biotechnology); anti-PML mouse monoclonal (sc-966; Santa Cruz Biotechnology); anti-Coilin mouse monoclonal (ab11822; Abcam); anti-BMI1 rabbit monoclonal (6964; Cell Signaling Technology); anti-SC35 mouse monoclonal (ab11826; Abcam); anti-CENPC guinea pig polyclonal (PD030; MBL, Medical & Biological Laboratories); and anti-phospho-Histone H2A.X (Ser139) mouse monoclonal (05-636; Millipore). The following secondary antibodies were purchased from Thermo Fisher Scientific: Alexa Fluor 488-conjugated anti-mouse (A-11029), anti-rabbit (A-11036); Alexa Fluor 568-conjugated anti-mouse (A-11031), anti-rabbit (A-11036), and anti-guinea pig (A-11075); and Alexa Fluor 647-conjugated anti-mouse (A-21236), and anti-rabbit (A-21245). Antibodies were diluted in PBS with 0.1% FBS. Samples were incubated with antibodies for 1 h at room temperature. After incubation, cells were treated with DAPI (Thermo Fisher Scientific) for 15 min and coverslipped (Matsunami) with Slow-Fade Gold (Thermo Fisher Scientific). For time-lapse experiments, the medium was replaced with Phenol red-free D-MEM/F12 (Thermo Fisher Scientific) supplemented with 10% FBS. Cells were incubated for 1 h, transferred to an incubator microscope (described above), maintained at 37 °C in 5% CO₂, and imaged for 1 to 3 h.

Data processing of microscopy images

All image analysis was performed using the Metamorph software (Molecular Devices). For quantification of proteasome foci, maximum projections of 16 z-stack images (0.2 μm apart) were manually segmented with the nucleus (identified by DAPI staining), and foci number and diameter were analysed using Transfluor. For quantification of proteasome foci in time-lapse imaging analysis, the diameter was calculated from the average foci area per cell, and the median per view field was averaged. To generate 3D images, we took 30 optical sections spaced 0.06 μm apart. The image view of the 3D point (xy, yz, xz) was given by projective transformation. For quantification of circularity of proteasome foci, images were processed through the close–open filter and analysed by integrated morphometric analysis. Circularity was calculated according to the following formula: circularity = $4\pi(\text{area}/\text{perimeter}^2)$. For quantification of RPL29 condensates, a maximum projection of three z-stacks (0.2 μm apart) was manually segmented with the nucleoplasm, and the number of foci per cell and the diameter were analysed using integrated morphometric analysis. The Pearson correlation coefficient (*r*) was calculated from a scatter plot of the fluorescence intensities of two proteins.

FRAP analysis

In-cell FRAP analysis was performed using a Leica TCS SP8 laser-scanning microscope equipped with a HC PL APO 100×/1.40 NA oil CS2. Intracellular assemblies were bleached in a circular 1-μm² region of interest using a 3.25-s pulse of the 405-nm laser line at full power. Recovery was monitored every 0.65 s for 400 frames. In vitro FRAP was performed using an Olympus Fluoview FV3000 equipped with a UAPON 100× OTIRF 1.49 NA (Olympus). Droplet assemblies were bleached in a circular 0.4-μm² region of interest using a 0.5-s pulse of the 488-nm laser line at full power. Recovery was monitored every 0.5 s for 180 frames. Recovery curves were analysed using Metamorph. Plotting and curve fitting were carried out in GraphPad Prism 7 (Graphpad Software).

Transmission electron microscopy

Samples were fixed with 2% paraformaldehyde and 2% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4) at 37 °C for 30 min, and then at 4 °C for 30 min. Afterwards, they were fixed with 2% glutaraldehyde 0.1 M phosphate buffer (pH 7.4) at 4 °C overnight. The cells were then washed with the same buffer and post-fixed with 2% osmium tetroxide (OsO₄) in the same buffer for 1 h. The samples were then dehydrated for 1 h in a graded ethanol solution and embedded in a resin (Quetol-812; Nissin EM) for 2 days, and polymerized at 60 °C for 48 h. Ultrathin sections (70-nm thickness) were made with a diamond knife on a Leica Ultracut UCT ultramicrotome (Leica), and then mounted on copper grids. Sections were stained with 2% uranyl acetate and lead stain solution (Merck) and visualized on JEM-1400Plus electron microscope (JEOL) at an acceleration voltage of 100 kV. Digital images (3,296 × 2,472 pixels) were acquired with an EM-14830RUBY2 camera (JEOL).

Correlative light electron microscopy

Image characteristics (pixel size, lateral resolution, and axial resolution) of fluorescence emission and electron density were merged using ec-CLEM on the open-source software platform Icy⁴⁰ (Institut Pasteur). For alignment, this algorithm relies on manual identification of matching landmarks in the cell in single-section images, and it operates with an accuracy determined by the degree of sample distortion caused by the intermediate sample preparation step.

Immunoprecipitation

Cells were collected and lysed in buffer A (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10% glycerol, 10 mM iodoacetamide) containing 0.2% NP-40 and complete protease inhibitor cocktail (Roche, EDTA-free). After standing on ice for 30 min the lysate was sonicated on a Handy Sonic

(Tomy Seiko), and then collected by centrifugation (20,000g for 15 min at 4 °C). The protein concentration was determined using the BCA protein assay kit (Thermo Fisher Scientific). For anti-Flag immunoprecipitation, anti-DDDDK-tag mAb-Magnetic beads (Medical and Biological Laboratories), M185-11) were used to precipitate Flag-tagged protein complex from 1 mg of cell lysate by incubating for 1 h at 4 °C. After five washes with buffer A containing 0.2% NP-40, the proteins were eluted for 10 min at 70 °C in 1× NuPAGE LDS sample buffer.

SDS–PAGE and immunoblotting

Cells were lysed with buffer A with complete protease inhibitor cocktail (EDTA free) (Roche) for 30 min on ice, and then centrifuged at 12,000g for 5 min at 4 °C. Supernatants were collected, and protein concentration was measured using the BCA protein assay kit (Thermo Fisher Scientific). Cell lysates were boiled in 1× LDS NuPAGE sample buffer for 10 min, and then electrophoresed on 4–12% NuPAGE Bis-Tris gels (Thermo Fisher Scientific). Proteins were transferred to polyvinylidene difluoride membranes (Millipore). The membranes were blocked for 30 min in 5% non-fat milk, and then incubated 1 h at room temperature with primary antibodies. The primary antibodies used were as follows: anti-Flag mouse monoclonal (A8592; Sigma-Aldrich); anti-multi-ubiquitin rabbit polyclonal (Z0458; Dako); anti-PSMB2 rabbit polyclonal (BML-PW8890; Enzo Life Sciences); anti-PSMD6 rabbit polyclonal (BML-PW8225; Enzo Life Sciences); Lys48-Specific anti-ubiquitin rabbit monoclonal (05-1307; clone Apu2; Millipore); anti-RAD23B rabbit monoclonal (13525; Cell Signaling Technology); anti-UBE3A rabbit polyclonal (10344-1-AP; Protein Tech); anti-cleaved caspase-3 (Asp175) rabbit polyclonal (9661; Cell Signaling Technology); and anti-β-actin pAb-HRP-Direct (PD030; MBL). After extensive washing with TBST, the membranes were immunoblotted with secondary antibodies for 30 min at room temperature. The following secondary antibodies were purchased from Jackson ImmunoResearch Laboratories: HRP-conjugated goat anti-rabbit Ig and HRP-conjugated goat anti-mouse Ig. After washing several times with TBST, blots were developed using ECL Prime Western Blotting Detection Reagent (GE Healthcare) and analysed on an ImageQuant LAS4000 (GE Healthcare).

Northern blotting

Total RNA was isolated with TRIzol reagent (Thermo Fisher Scientific). Total RNA (4 µg per lane) was resolved on a 1.2% formaldehyde agarose denaturing gel in 1× TT (30 mM Tricine and 30 mM triethanolamine) buffer by electrophoresis for 120 min at 200 V, followed by capillary transfer to Hybond-N+ membrane (GE Healthcare) with 20× SSC (3 M NaCl and 300 mM trisodium citrate dihydrate) for 18 h. After transfer, RNA was UV cross-linked to the membrane using a CL-1000 ultraviolet crosslinker (UVP) at 120 mJ cm². Membranes were pre-hybridized with DIG Easy Hyb Granules (Roche) dissolved in double-distilled water (hybridization buffer), for 1 h in a hybridization oven at 50 °C. DIG-labelled DNA probes were added to the buffer and incubated for 20 h. Membranes were washed with non-stringent buffer (2× SSC and 0.1% SDS) for 15 min at 50 °C, followed by stringent buffer (0.1× SSC and 0.1% SDS) twice for 15 min each. The membrane was then washed with 1× MA buffer (100 mM maleic acid and 150 mM NaCl, pH 7.0) for 10 min at room temperature, and incubated with Blocking Reagent (Roche) for 30 min. Anti-digoxigenin-AP, Fab fragments (Roche) was added to the blocking reagent, and further incubated for 1 h. After that, the membranes were washed three times for 10 min each with 1× MA buffer containing 0.3% Tween-20, and then equilibrated in buffer-A (100 mM Tris-HCl and 100 mM NaCl, pH 9.5). To detect RNA, the membranes were incubated with CDP-star (Roche) for 10 min, and then chemiluminescence was visualized on a LAS-4000 mini (GE healthcare). DNA probes were labelled with DIG at the 5' end; probe sequences were as follows: 5'ETS; 5'-CGGA GGCCCAACCTCTCCGACGACAGGTCGCCAGAGGACAGCGTGTCAGC-3'; ITS1; 5'-CCTCGCCCTCCGGGCTCCGGGCTCCGTTAATGATC-3'; and ITS2; 5'-CTGCGAGGGAACCCCCAGCCGCGCA-3'.

Purification of recombinant proteins

Codon-optimized human RAD23B or RAD23A cDNA (Eurofins Genomics) was subcloned into *Escherichia coli* expression vector pGEX6P1 (GE Healthcare), in which the Cys sequence (TGC) was inserted upstream of the BamHI site for Cy labelling. RAD23B mutants, the UBL domain (Δ1–79), or two UBA domains (Δ180–238/Δ353–409), were generated by inverted PCR. Human ubiquitin with a N-terminal Met–Cys was inserted into vector pET26b (Novagen). GST–Cys–RAD23 proteins or Met–Cys-ubiquitin were expressed in *E. coli* BL21 (DE3) cells by induction with 0.2 mM IPTG at 28 °C for 3 h. GST-tagged proteins were adsorbed onto glutathione–Sephrose 4B (GE Healthcare) in buffer B (50 mM HEPES-KOH, pH 7.5, 100 mM NaCl and 10% glycerol) containing 0.1% Triton X-100. After washing with the same buffer, the beads were incubated with PreScission protease (GE Healthcare) to cleave the GST-tag from GST–Cys–RAD23 proteins, and the Cys–RAD23 proteins were recovered. Cells expressing Met–Cys–Ubiquitin were suspended in 50 mM HEPES-KOH, pH 7.5 and 100 mM NaCl and lysed by sonication. After centrifugation at 20,000g for 30 min, the supernatant was recovered and boiled at 80 °C for 5 min. The lysate was further clarified by centrifugation (20,000g for 20 min) to obtain Met–Cys–Ubiquitin. Cy3 or Cy5 labelling was performed using Cy Maleimide Mono reactive dye (GE Healthcare). After Cy labelling, Cy5–ubiquitin and Cy3–RAD23 proteins were separated from free dye on a PD-10 column (GE Healthcare). Unanchored K48-linked ubiquitin chains and K63-linked ubiquitin chains were prepared using UBE2K/E2-25K and Ubc13/Mms2, respectively, as previously reported⁴¹. To obtain Cy5-labelled ubiquitin chains with defined lengths, we further purified the ubiquitin chains by Superdex 75 10/300 GL size exclusion chromatography (GE Healthcare).

In vitro phase-separation assay

Droplet formation of the purified protein was monitored by fluorescence and differential interference contrast (DIC) microscopy using an Olympus Fluoview FV3000 (described above). Unless otherwise noted, 20 µM Cy5–ubiquitin and 20 µM Cy3–RAD23 proteins were incubated at room temperature in 50 mM Tris-HCl (pH 7.5), 3% PEG (Sigma-Aldrich) and 200 mM NaCl. The mixture in a total solution volume of 10 µl was placed in a glass-bottom dish (MatTek) or Sensoplate glass-bottom 384-well plate (Greiner Bio-One) coated with 0.1% PVA (Sigma-Aldrich).

Phase diagram mapping

After mixing RAD23B and ubiquitin proteins as described above, the mixture was placed on Sensoplate glass-bottom 384 well plates (Greiner) coated with 0.1% PVA (Sigma-Aldrich) and incubated for 90 min at room temperature. The presence or absence of droplets was scored by Cy3 and Cy5 fluorescence in each well at 100× magnification using an Olympus Fluoview FV3000. For fluorescence observation, the maximum projection of three z-stacks from the bottom of the plate was processed, and the scale was set in the range from 0 to 4,095. Two wells were prepared per experiment, and if fluorescence was recorded at three or more areas in each well, it was regarded as a droplet. The experiment was performed twice.

Quantification of RAD23B and RAD23A in HCT116 cells

Cellular abundance of RAD23A and RAD23B were determined by quantitative western blotting using recombinant protein as a standard. After washing twice with PBS, HCT116 cells were lysed by sonication in 2% SDS, 20 mM HEPES-Na (pH 7.5) and 1 mM EDTA. Proteins were subjected to western blotting with anti-RAD23A antibody (24555; Cell Signaling Technology) or anti-RAD23B antibody (13525; Cell Signaling Technology). Copy number per cell and concentration were calculated, assuming that total protein per cell was approximately 200 pg and that cell volume was approximately 1 pl, as described previously^{26,42}.

Propidium iodide exclusion assay and measurement of caspase-3 activity

Cell death under hyperosmotic stress was quantified by propidium iodide (PI) exclusion. Immediately after washing with PBS, PI (Biovision) was added to the cells at a final concentration of 50 $\mu\text{g ml}^{-1}$ and incubated at room temperature for 15 min, after which excess PI was washed off with PBS. At least two random fields were imaged using Olympus Fluoview FV3000 and processed with the MetaMorph software (Molecular device). Cell death was quantified by calculating the number of PI-labelled cells as a percentage of the total number of cells observed by DIC. Early apoptotic activity was assessed using the cleaved-caspase-3 (Asp175) antibody purchased from Cell Signaling Technology (9661). The experiment was repeated three times.

Statistical analysis

Statistical analysis was performed using GraphPad Prism 7. All details on statistics have been indicated in figure legends and Source Data. Sample distribution was assessed with the D'Agostino-Pearson normality test or the Kolmogorov-Smirnov test depending on sample size. Unpaired two-tailed Student's *t*-test was used to determine statistical significance when comparing two independent groups with normal distribution and no significant difference in their s.d. When comparing two independent groups in which at least one did not fit the normality criteria, unpaired two-tailed Mann-Whitney *U*-test was used. For analysis of statistical significance in comparisons involving more than two groups with normal distribution, ordinary one-way ANOVA with Dunnett's (when comparing the mean of each group with the mean of a control group) multiple comparisons test was used. For multiple comparisons in which at least one group did not comply with normal distributed data, Kruskal-Wallis and Dunn's post hoc tests were used. In all cases, statistical significance was assessed with a 95% confidence interval; therefore, $P < 0.05$ was considered significant.

Purification of ubiquitylated proteins and sample preparation for mass-spectrometry analysis

HCT116 cells (1×10^7) were treated with 0.2 M sucrose for the indicated periods of time (0, 15, 30 min) before collection. The cells were lysed in a urea-containing lysis buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10% glycerol and 6 M Urea) supplemented with 10 μM bortezomib (LC Laboratories), 10 mM iodoacetamide (GE Healthcare) and 1 \times Complete protease inhibitor cocktail (Roche, EDTA free), and sonicated on a Handy Sonic (Tomy Seiko). The lysate was then clarified by centrifugation (20,000g for 20 min), and the protein concentration was determined using the BCA protein assay kit (Thermo Fisher Scientific). To capture ubiquitylated proteins, we used a biotinylated trypsin-resistant tandem ubiquitin-binding entity (^{BIO}TR-TUBE)⁴¹. ^{BIO}TR-TUBE (10 μg) and 5 mg of cell lysate were incubated for 1 h at 4 °C and incubated for an additional 1 h at 4 °C with 50 μl of Dynabeads MyOne streptavidin C1 (Thermo Fisher Scientific). After two washes with lysis buffer containing 6 M urea and two washes with urea-free lysis buffer, the beads were suspended in 50 μl of 50 mM ammonium bicarbonate (AMBC) and 0.1% RapiGest SF (Waters). Proteins on beads were reduced in 5 mM TCEP for 30 min at 50 °C and then alkylated with 10 mM methylmethanethiosulfonate for 10 min at room temperature. The alkylated proteins were digested overnight at 37 °C with 1 μg of trypsin Gold (Promega). After tryptic digestion, samples were acidified to pH 2 with trifluoroacetic acid (TFA) and desalted by solid-phase extraction using GL-Tip SDB and GL-Tip GC (GL Sciences). Ubiquitinated peptides were enriched using the PTMScan ubiquitin Remnant Motif (K- ϵ -GlyGly) Kit (Cell Signaling Technology)⁴³. After three washes with 200 mM triethanolamine (pH 8.2), antibody-coupled beads were cross-linked with 20 mM dimethyl pimelimidate in 200 mM triethanolamine (pH 8.2) for 1 h at room temperature. The beads were then washed three times with 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, and stored at 4 °C until use. Trypsinized

peptides prepared as described above were suspended in 0.2 ml IAP buffer (50 mM MOPS, pH 7.2, 10 mM sodium phosphate and 50 mM NaCl) and incubated with 10 μl of antibody-coupled beads for 2 h at 4 °C. After two washes with 250 μl of IAP buffer and three washes with 250 μl of Milli-Q water (Millipore), di-Gly peptides were eluted with 3 \times 20 μl of 0.15% TFA. The eluted peptides were desalted to 0.1% TFA using GL-Tip SDB and GL-Tip GC (GL Sciences), and analysed by mass spectrometry.

Purification of proteasome-interacting proteins for mass-spectrometry analysis

After addition of 0.2 M sucrose and incubation for 30 min, HCT116 PSMD6-eGFP-3 \times Flag^{KI/KI} cells were cross-linked with 0.2% formaldehyde (Thermo Scientific) for 10 min, after which the reaction was quenched with 0.25 M glycine. After two washes with PBS, the cells were lysed with lysis buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10% glycerol, 4 mM ATP, 10 mM MgCl₂, 0.2% NP-40, 1 \times Protein Complete Inhibitor-EDTA, 1 mM bortezomib, 1 mM iodoacetamide) with sonication. Cleared lysates (2 mg) were incubated with 10 μl of anti-DDDDK antibody-coupled agarose (MBL) for 1 h at 4 °C. After the beads were washed five times with lysis buffer, proteins were eluted with 3 \times Flag peptide (Sigma). The eluted proteins were mixed with 3 \times LDS NuPAGE sample buffer and incubated for 30 min at 95 °C for uncrosslinking. Proteins were then electrophoresed 1 cm by 4–12% NuPAGE. The gels were stained with Bio-Safe Coomassie Stain (Bio-Rad). After washing extensively with Milli-Q water, the gels were excized, cut into 1-mm³ pieces, destained twice for 1 h each with 1 ml 50 mM AMBC/50% acetonitrile (ACN), with agitation, and then dehydrated with 100% ACN. Trypsin solution (Promega, 20 ng μl^{-1} in 50 mM AMBC/5% ACN) was added to the gel pieces, and the samples were incubated at 37 °C overnight. Digests were extracted with 100 μl of 50% ACN/0.1% TFA. The extracted peptides were concentrated using a speed-vac and resuspended in 0.1% TFA.

Mass-spectrometry analysis

Shotgun mass-spectrometry analysis was performed essentially as described^{41,44}. An Easy nLC 1000 system (Thermo Fisher Scientific) was connected online to a Q Exactive mass spectrometer (Thermo Fisher Scientific) with a nanoelectrospray ion source. The mobile phases were 0.1% formic acid (FA) in water (solvent A) and 0.1% FA in 100% ACN (solvent B). Peptides were directly loaded onto a C18 analytical column (ReproSil-Pur 3 μm , 75- μm inner diameter and 12-cm length, Nikkyo Technos) and separated using a 150-min three-step gradient (0–40% solvent B for 120 min, 40–100% for 20 min and 100% for 10 min) at a constant flow rate of 300 nL min⁻¹. For ionization, liquid junction voltage was 1.8 kV and capillary temperature was 250 °C. The Q Exactive was operated by the Xcalibur software 2.2 (Thermo Fisher Scientific) in data-dependent MS/MS mode, and the top 10 most intense ions with charge state +2 to +5 were selected with an isolation window of $m/z = 2.0$ and fragmented by higher-energy collisional dissociation with a normalized collision energy of 25. Resolution and automatic gain control target for the survey scans were set to 70,000 and 3E6, respectively. Ions selected for MS/MS were dynamically excluded for 5 s (for diGly peptide identification) or 30 s (for binding protein identification). The data were analysed using the Mascot search program (Matrix Science) in Proteome Discoverer 1.3 (Thermo Fisher Scientific). Maximum missed cleavage site of trypsin was set to two. Oxidation (Met), GlyGly (Lys), phosphorylation (Ser, Thr, Tyr) and pyroglutamate conversion (Gln) were selected as variable modifications. For the diGly proteome, Cys-methylthio modification was set as a static modification for database searches. Peptide identification was filtered at a false discovery rate <0.01.

Cryo-EM

For sample preparation, 25,000 cells were seeded on EM grids (R2/1, Au 200 mesh grid, Quantifoil Micro Tools) in a 35-mm dish and cultured in

DMEM supplemented with penicillin–streptomycin, 1% non-essential amino acids and 10% fetal bovine serum overnight. Thirty minutes after sucrose stimulation by replacing the medium with 0.2 M sucrose, the grids were blotted for 10 s using filter paper and vitrified by plunge-freezing into a liquid ethane/propane mixture using a Vitrobot Mark IV (FEI). Cryo-focused ion beam (FIB) microscopy cryo-electron tomographic data collection was performed as described in detail before⁴⁵. In brief, electron microscopy grids were mounted onto modified Autogrids sample carriers⁴⁶, and then transferred into a dual-beam (FIB/SEM) microscope (Quanta 3D FEG, FEI) using a cryo-transfer system (PP3000T, Quorum). For the whole procedure, samples were kept at a constant temperature of -180°C . Thin lamellae with a thickness of about 200 nm were prepared in the nuclear region of the cells using a Ga^{2+} ion beam at 30 kV under a 20° stage tilt angle. The milling progress was monitored by SEM imaging at 5 kV. The lamellas were transferred to an FEI Titan Krios transmission electron microscope equipped with a Gatan post-column energy filter and Gatan K2 Summit direct detector for tomographic data collection. Tilt series were recorded from -50° to $+70^{\circ}$ with an increment of 2° using SerialEM software⁴⁷ at a nominal magnification of $42,000\times$, resulting in a pixel size of 3.42 \AA at the specimen level. On average, six frames were collected for each image under counting mode, resulting in a total dose of $110\text{ e}^{-}\text{ \AA}^{-2}$ per tilt series. For image processing, the MATLAB (Mathworks) TOM toolbox⁴⁸ was used as a general platform for image processing. K2 frames were aligned using in-house software (K2Align) based on previous work⁴⁹. K2Align code is available at GitHub (<https://github.com/dtegunov/k2align>). Using the IMOD software package, tilt series were first aligned using fiducial-less tracking, and then tomograms were reconstructed by weighted back-projection of the resulting aligned tilt series⁵⁰. To identify proteasomes in the tomograms, a mirrored single-capped proteasome structure (EMD-3916) was filtered to 60 \AA as an initial template for template matching in the twice-binned tomograms (13.68 \AA^3 per voxel) using PyTom⁵¹. The resulting subtomograms were cropped out, CTF-corrected and classified using RELION⁵². The resulting structure clearly showed a 26S proteasome complex with correct handedness. However, owing to limitations on the number of subtomograms, we could not separate more conformational information.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw cryo-ET data have been deposited to the Electron Microscopy Data Bank under accession code EMD-10494. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD01637 and PXD016369. The uncropped blots and gels are provided in Supplementary Fig. 1. Source Data for Figs. 1–4 and Extended Data Figs. 1, 3–9 are provided with the paper.

Code availability

K2Align code is available at GitHub (<https://github.com/dtegunov/k2align>).

31. Berkers, C. R. et al. Profiling proteasome activity in tissue with fluorescent probes. *Mol. Pharm.* **4**, 739–748 (2007).
32. Kroschwald, S., Maharana, S. & Simon, A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* <https://doi.org/10.19185/matters.201702000010> (2017).
33. D'Arcy, P. et al. Inhibition of proteasome deubiquitinating activity as a new cancer therapy. *Nat. Med.* **17**, 1636–1640 (2011).
34. Hyer, M. L. et al. A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment. *Nat. Med.* **24**, 186–193 (2018).
35. Magnaghi, P. et al. Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nat. Chem. Biol.* **9**, 548–556 (2013).
36. Sakuma, T. et al. Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci. Rep.* **3**, 3379 (2013).
37. Sanjana, N. E. et al. A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* **7**, 171–192 (2012).
38. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
39. Kumatori, A. et al. Abnormally high expression of proteasomes in human leukemic cells. *Proc. Natl Acad. Sci. USA* **87**, 7071–7075 (1990).
40. Paul-Gilloteaux, P. et al. eC-CLEM: flexible multidimensional registration software for correlative microscopies. *Nat. Methods* **14**, 102–103 (2017).
41. Tsuchiya, H. et al. In vivo ubiquitin linkage-type analysis reveals that the Cdc48-Rad23/Dsk2 axis contributes to K48-linked chain specificity of the proteasome. *Mol. Cell* **66**, 488–502 (2017).
42. Fujioka, A. et al. Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *J. Biol. Chem.* **281**, 8917–8926 (2006).
43. Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* **8**, 1950–1960 (2013).
44. Yoshida, Y. et al. A comprehensive method for detecting ubiquitinated substrates using TR-TUBE. *Proc. Natl Acad. Sci. USA* **112**, 4630–4635 (2015).
45. Guo, Q. et al. In situ structure of neuronal C9orf72 Poly-GA aggregates reveals proteasome recruitment. *Cell* **172**, 696–705 (2018).
46. Rigort, A. et al. Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. *Proc. Natl Acad. Sci. USA* **109**, 4449–4454 (2012).
47. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
48. Nickell, S. et al. TOM software toolbox: acquisition and analysis for electron tomography. *J. Struct. Biol.* **149**, 227–234 (2005).
49. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
50. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
51. Hrabe, T. et al. PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* **178**, 177–188 (2012).
52. Bharat, T. A. & Scheres, S. H. Resolving macromolecular structures from electron cryotomography data using subtomogram averaging in RELION. *Nat. Protoc.* **11**, 2054–2065 (2016).

Acknowledgements We thank C.-G. Pack and M.-K. Jung for initial electron microscopy analysis; S. Fukai, K. Iwai, T. Yamamoto and F. Zhang for reagents; N. Noda for helpful discussion; and J. Horiuchi for critical reading of the manuscript. This research was supported by AMED under grant numbers JP18gm1110003 (S.M. and Y.S.) and JP19gm1110010 (T.I.); MEXT/JSPS KAKENHI grant numbers JP18K19352 (S.Y.), JP18K14913 (H.T.), JP26293018 (Y.S.), JP18H03993 (Y.S.), JP18H05498 (F.O. and Y.S.), JP18H03977 (T.I.), JP19H05281 (T.I.) and JP21000012 (K.T.); the Takeda Science Foundation (Y.S., K.T. and T.I.); the Uehara Memorial Foundation (Y.S.); the European Commission, FP7 GA ERC-2012-SyG_318987-ToPAG (Q.G., W.B. and R.F.-B.); and the DFG, EXC 2067/1- 390729940 (R.F.-B.).

Author contributions S.Y., K.T. and Y.S. designed most of the experiments; S.Y., A.K. and Y.S. generated the Ki/KO cell lines with assistance of S.M.; S.Y. and A.K. performed the fluorescent microscopy analyses; S.Y. processed microscopy images; H.T., A.K., N.A. and A.E. performed the immunoprecipitation and western blotting; H.T. performed the mass spectrometry proteomics analysis with assistance of A.E. and F.O.; H.T. and N.A. prepared recombinant proteins; S.Y. and Y.S. performed in vitro phase separation assay; Q.G., W.B. and R.F.-B. performed and analysed cryo-ET experiments; K.I. and T.I. performed northern blotting; S.Y., K.T. and Y.S. wrote the manuscript with input from all co-authors.

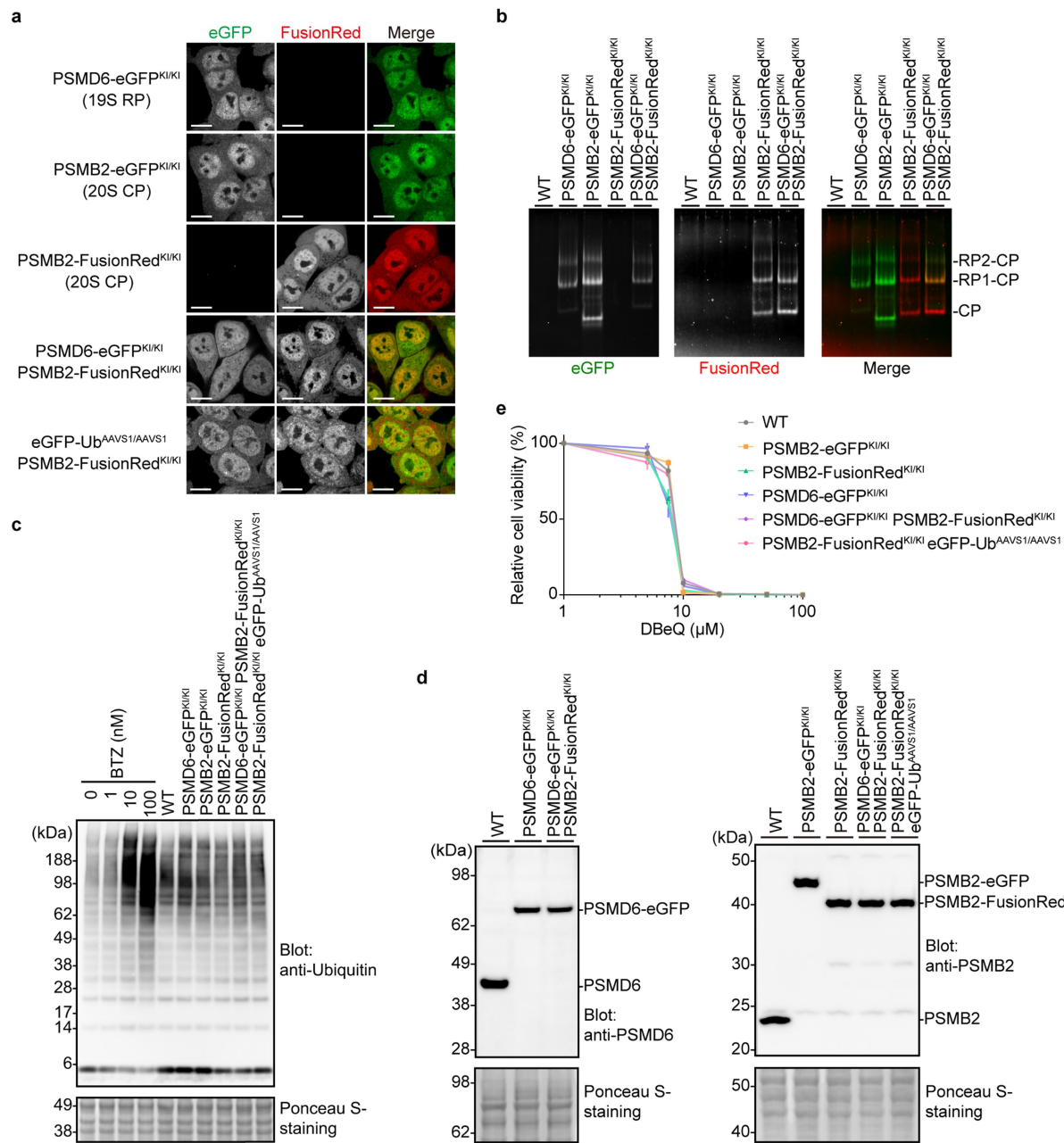
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1982-9>.

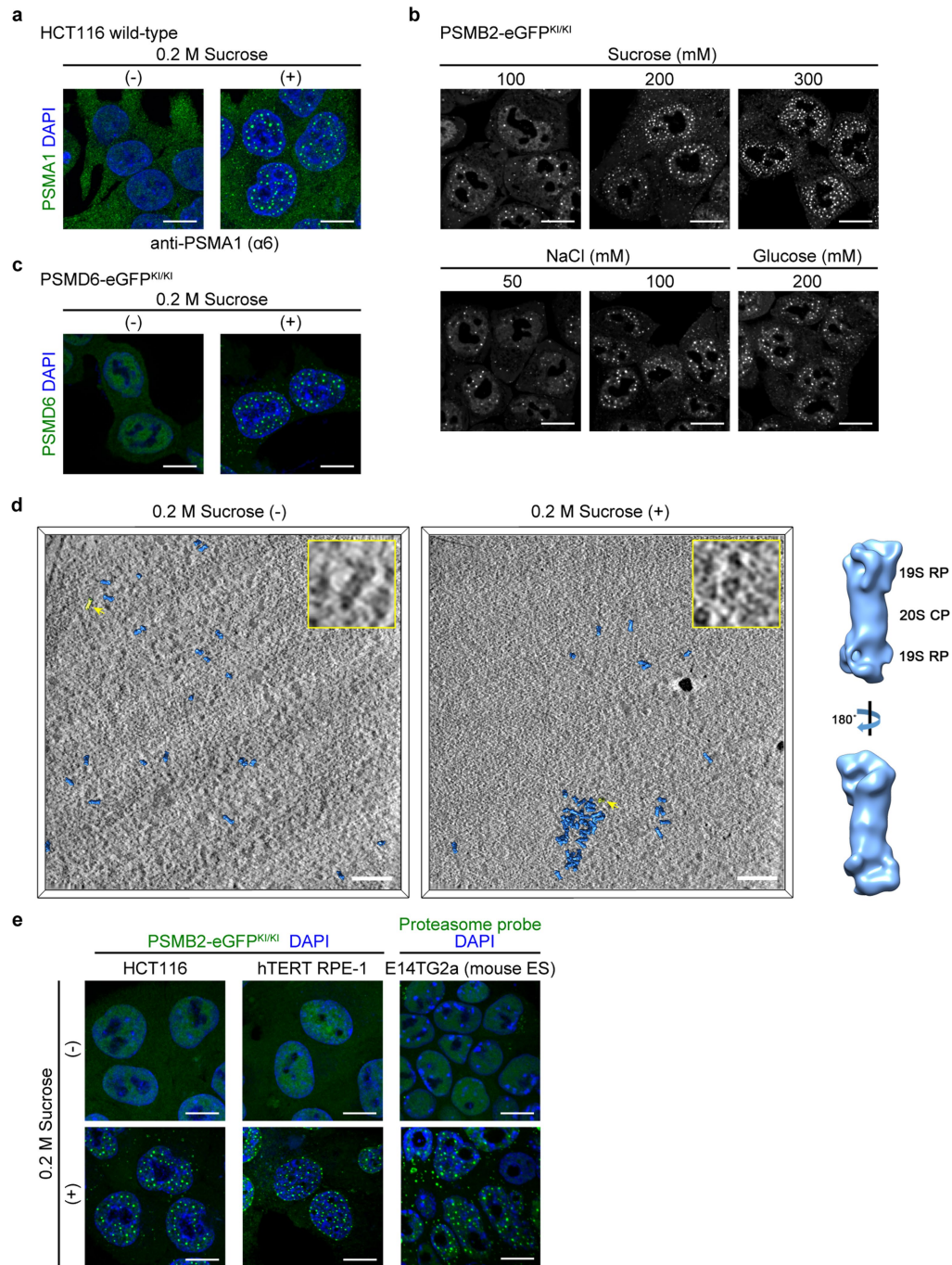
Correspondence and requests for materials should be addressed to K.T. or Y.S.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Characterization of KI cells stably expressing fluorescent protein-labelled proteasome subunit or ubiquitin. a, HCT116 cells stably expressing proteasome subunits (PSMB2 or PSMD6) or ubiquitin fused with either eGFP or FusionRed. Scale bars, 10 μm. **b**, Proteasome assembly state of cell lines indicated in **a**, as determined by native PAGE. **c**, **d**, Immunoblot analysis of the KI cells in **a** with antibodies against

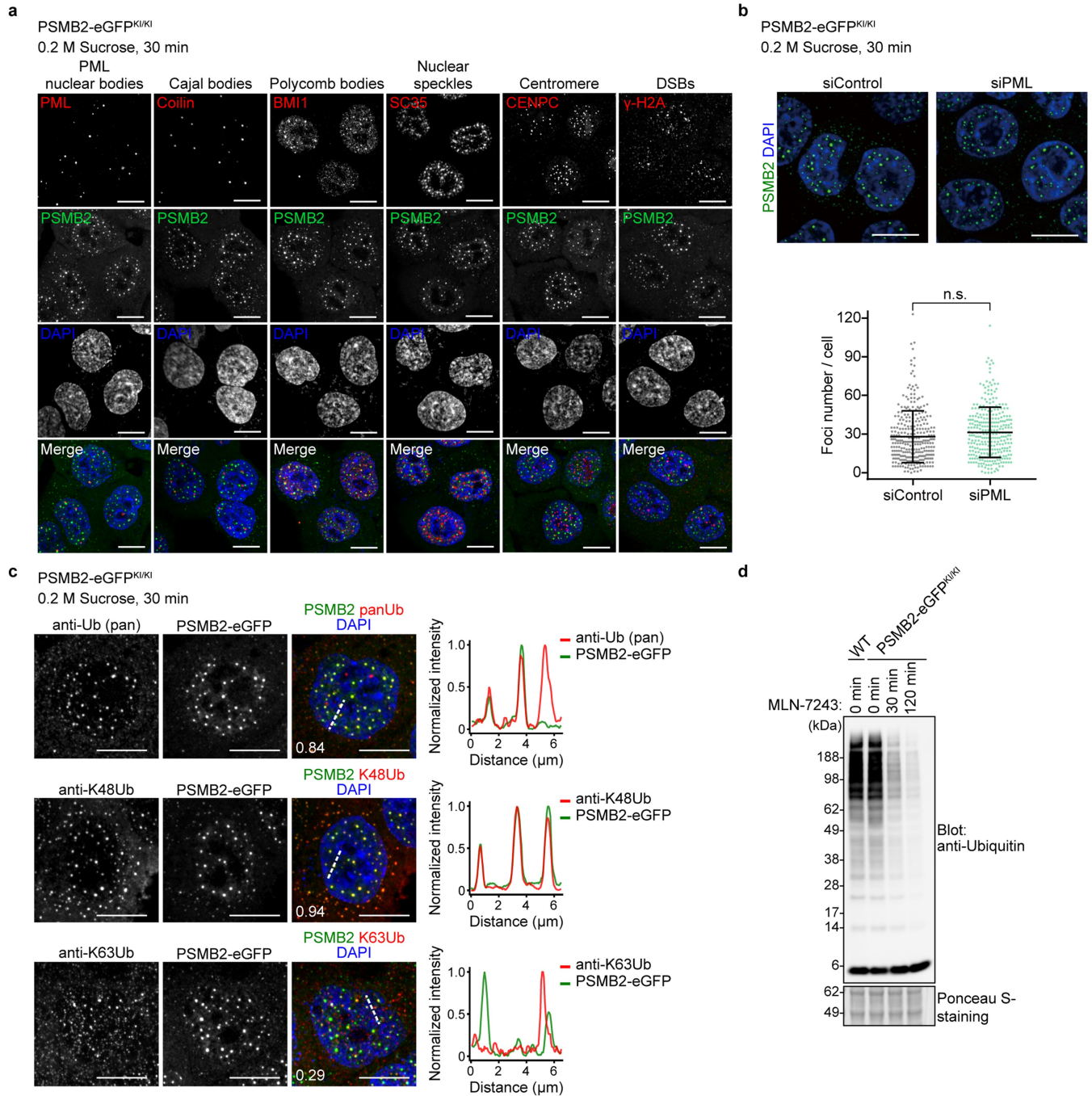
ubiquitin (**c**), PSMD6 or PSMB2 (**d**). Ponceau S staining was performed as a loading control. **e**, Cell viability of DBeQ-treated cells was determined using the CellTiter assay kit. The MTS assay was performed in triplicate. Data are mean ± s.d. (n = 3). In **a–d**, representative results from two independent experiments. Gel source data for **b–d** are shown in Supplementary Fig. 1.



Extended Data Fig. 2 | Proteasome foci formation occurs in various cell types following hyperosmotic stress.

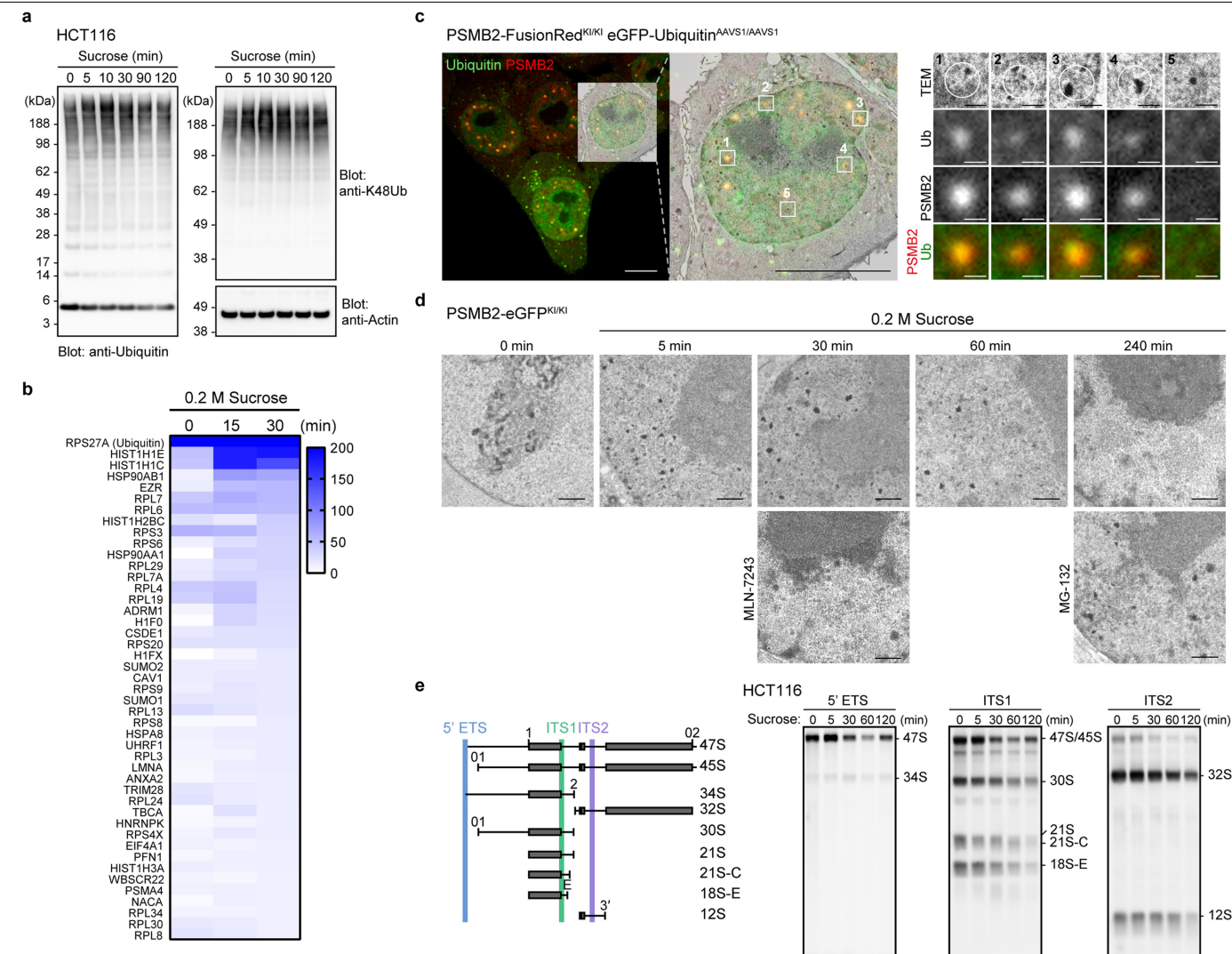
a, Localization of endogenous proteasome in HCT116 cells, with or without 0.2 M sucrose stimulation for 30 min, was observed using PSMA1 (α6) antibodies. Scale bars, 10 μm. **b**, PSMB2-eGFP^{K1/K1} cells were stimulated with sucrose, NaCl or glucose at the indicated concentrations for 5 min. Scale bars, 10 μm. **c**, PSMD6-eGFP^{K1/K1} cells were stimulated with 0.2 M sucrose for 30 min. Scale bars, 10 μm. **d**, Tomographic slices of the nuclear region in cells not stimulated (left) or stimulated with 0.2 M sucrose (right). For the whole tomogram, proteasomes were mapped to their original positions and orientations by template matching and sub-tomogram averaging. Higher-magnification tomographic slices of a representative proteasome (arrow; coloured in yellow) detected in the tomograms are shown

in the top right. In total, nine tomograms were collected from sucrose-stimulated cells, of which three exhibited proteasome clustering. Seven tomograms were collected from non-stimulated cells, none of which exhibited proteasome clustering. Three-dimensional reconstruction revealed a proteasome structure resolved by averaging 280 sub-tomograms from all tomograms obtained with and without sucrose stimulation. Scale bars, 0.2 μm. **e**, HCT116 cells and hTERT RPE-1 cells stably expressing PSMB2-eGFP and ES-E14TG2a cells from mouse embryos were stimulated with 0.2 M sucrose for 30 min. For ES-E14TG2a cells, endogenous proteasome activity was detected using a proteasome probe (Me4BodipyFL-Ahx3Leu3VS). Scale bars, 10 μm. Representative results from three (**a**) or two (**b–e**) independent experiments.



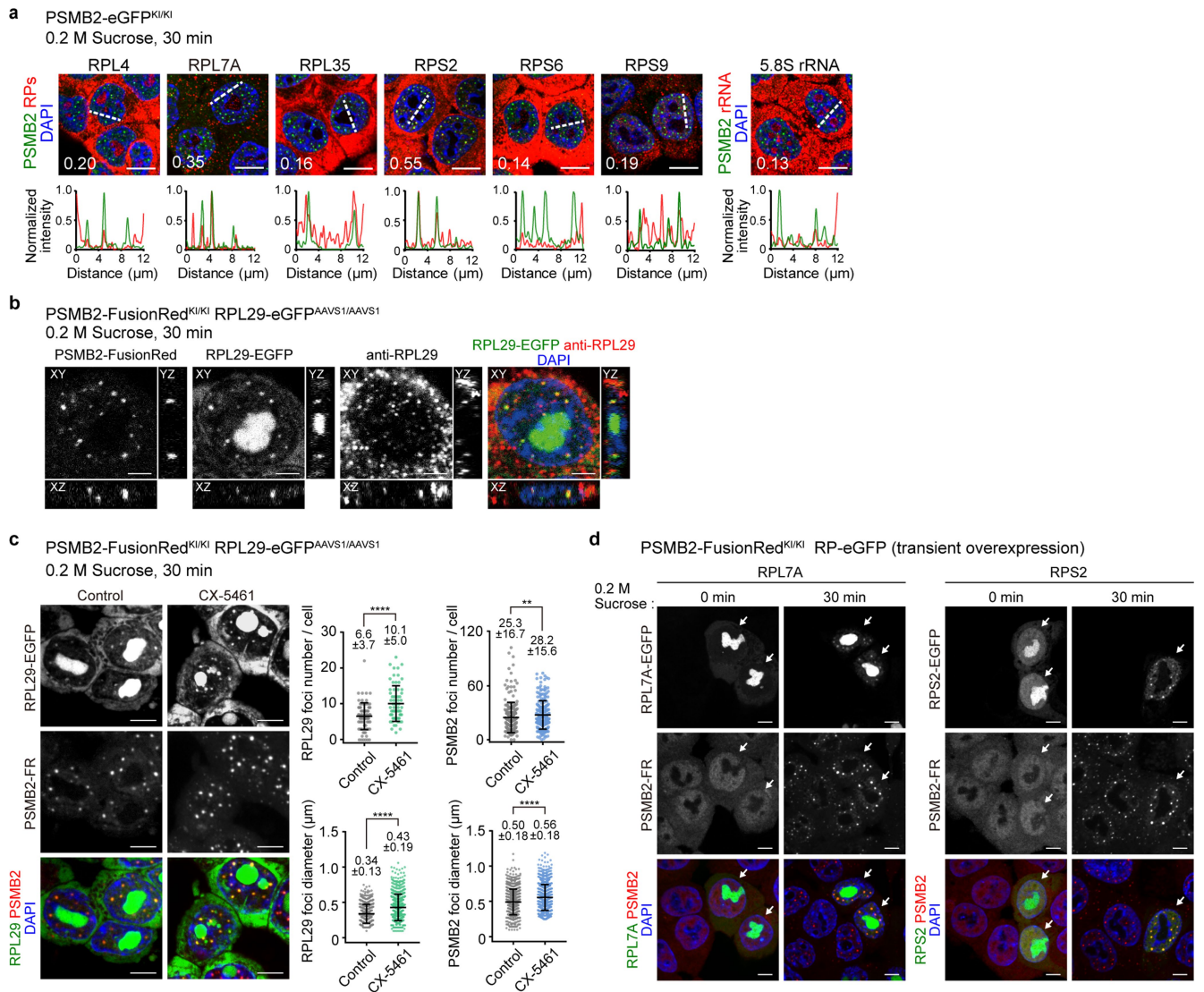
Extended Data Fig. 3 | Proteasome foci are distinct from known nuclear structures. **a**, Colocalization of proteasome foci induced by 0.2 M sucrose, as determined by staining for markers of various nuclear bodies. Scale bar, 10 μm. Representative images from two independent experiments. **b**, PSMB2-eGFP^{KI/KI} cells were transiently transfected with *PML* siRNA or control siRNA for 48 h, and then stimulated with 0.2 M sucrose for 30 min. The graph indicates the number of foci per cell (siControl, $n = 300$ cells; siPML, $n = 294$ cells). Data are mean \pm s.d., $P = 0.0541$ by two-tailed unpaired Student's *t*-test. Scale bars, 10 μm. **c**, PSMB2-eGFP^{KI/KI} cells were stimulated with 0.2 M sucrose for 30 min and were probed with pan-ubiquitin (anti-Ub, clone FK2), K48-ubiquitin

(anti-K48Ub) and K63-ubiquitin (anti-K63Ub) specific antibodies. Line profiling of representative sections of cells is indicated by white dashed lines. The mean value of the Pearson correlation coefficient in the nucleoplasm is shown in the image ($n = 10$ cells in two view fields). Scale bars, 10 μm. Representative images from three (anti-Ub, anti-K48) or two (anti-K63Ub) independent experiments. **d**, Wild-type or PSMB2-eGFP^{KI/KI} HCT116 cells were stimulated with the E1 inhibitor MLN-7243 (1 μM) for the indicated times and immunoblotted with ubiquitin antibody. Representative result from three independent experiments. For gel source data, see Supplementary Fig. 1.



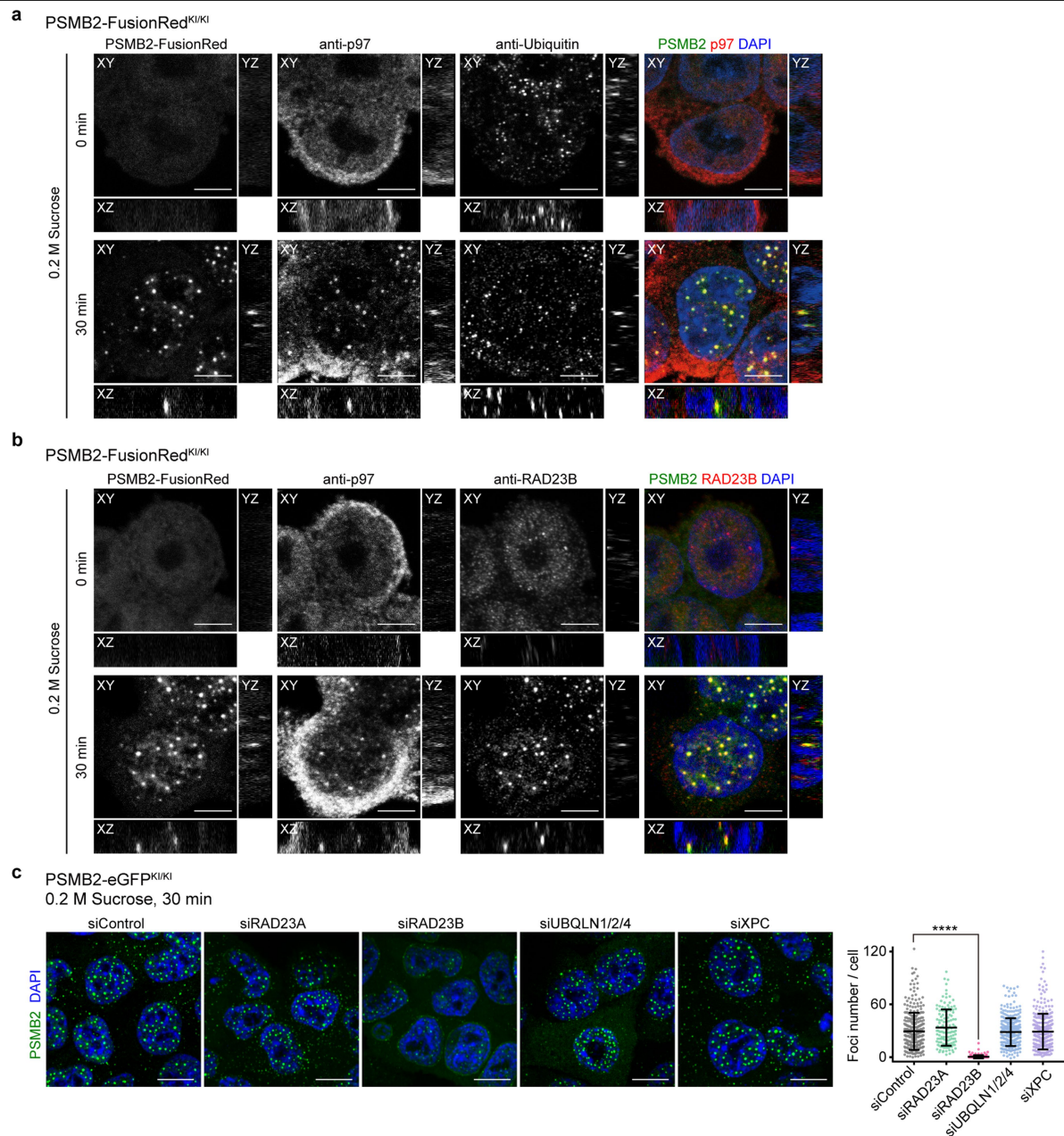
Extended Data Fig. 4 | Changes in ubiquitylation levels and intranuclear structure in response to hyperosmotic stress. **a**, HCT116 cells treated with 0.2 M sucrose for the indicated times were subjected to immunoblotting with indicated antibodies. Representative results from two (anti-K48) or three (anti-Ubiquitin) independent experiments. **b**, Mass spectrometry screening of ubiquitylated substrates by the double-concentration method using TR-TUBE and anti-diGly antibody. PSMB2-eGFP^{KI/KI} cells were stimulated with 0.2 M sucrose at the indicated times. The numbers of peptide spectrum matches for each identified protein are shown as a heat map. Similar results were obtained from two independent experiments. See Source Data for identified peptide list. **c**, Correlative light and electron microscopy (CLEM) analysis of PSMB2-FusionRed^{KI/KI}-eGFP-Ub^{AAVS1/AAVS1} cells. Scale bars, 10 μ m. Right, square enlarged

images on the left. The places where proteasome foci exist are circled in white. Scale bars, 0.5 μ m. Representative images from two independent experiments. **d**, Time course images of PSMB2-eGFP^{KI/KI} cells stimulated with 0.2 M sucrose, with or without pre-treatment with 1 μ M MLN-7243 or 50 μ M MG-132 for 1 h. Scale bars, 1 μ m. Two independent experiments were performed with 0 min, 30 min, 240 min and MG-132, with similar results. **e**, Schematic of northern blot probes and positions of the pre-rRNA processing intermediates discussed in this Article. RNA was extracted from HCT116 cells stimulated with 0.2 M sucrose for the indicated times and analysed by northern hybridization using probes specific for the 5' end and the middle of ITS1 and ITS2. Representative results from three independent experiments. Gel source data for **a** and **e** are shown in Supplementary Fig. 1.



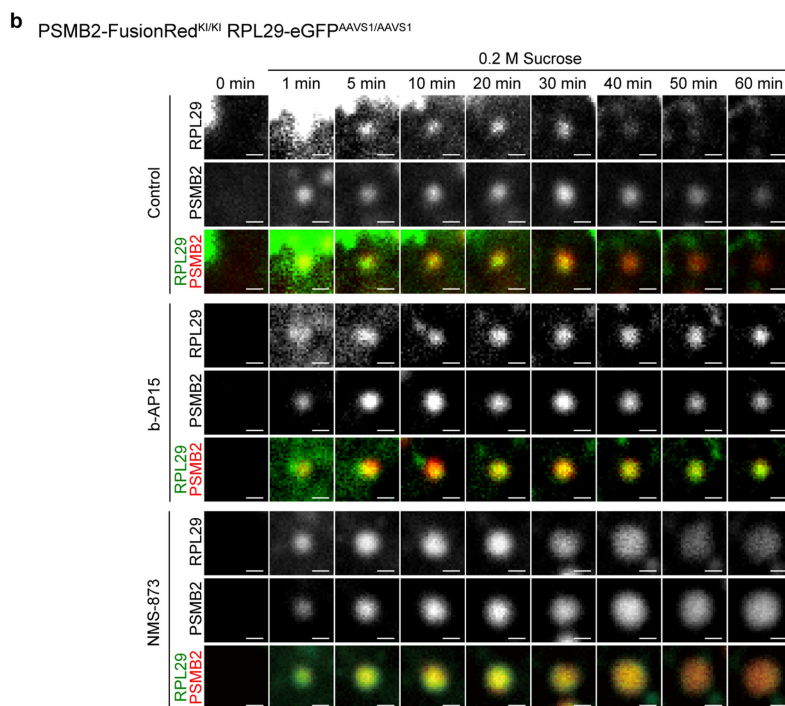
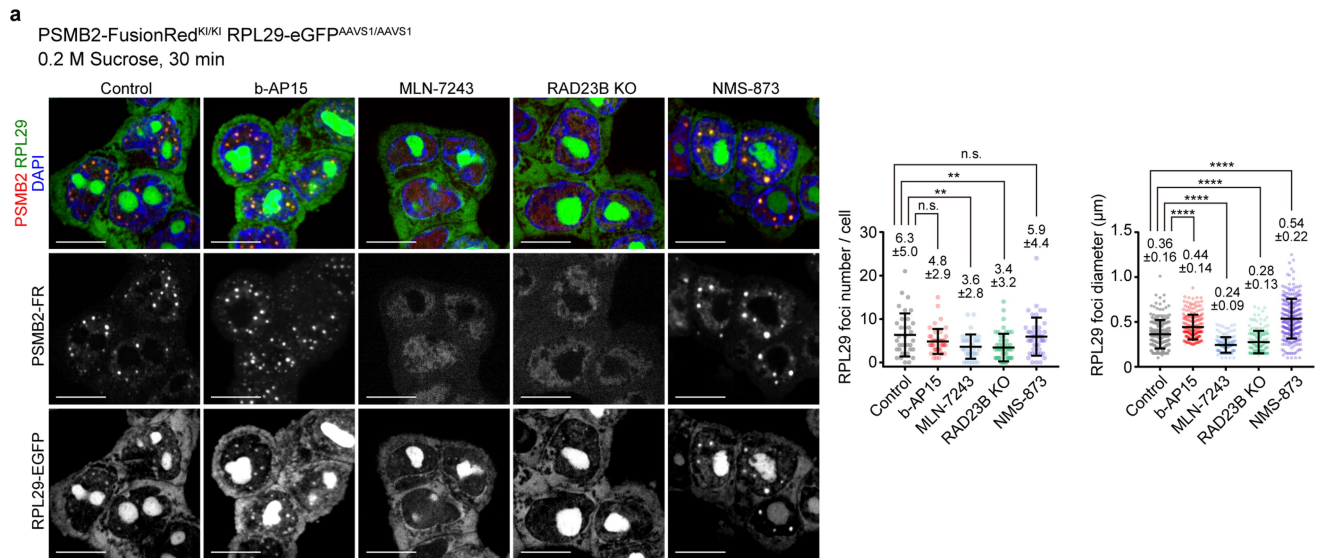
Extended Data Fig. 5 | Unassembled ribosomal proteins are present in proteasome foci. **a**, Colocalization of proteasome foci and endogenous ribosome components. PSMB2-eGFP^{KI/KI} cells were stimulated with 0.2 M sucrose for 30 min, and endogenous ribosome components were detected by specific antibodies. The mean value of the Pearson correlation coefficient in the nucleoplasm is shown in the image ($n = 10$ cells in two fields of view). Scale bars, 10 μ m. Each graph represents the normalized fluorescence distribution over white dashed lines of the cells. Representative images from two (RPL4, RPL7A, RPL35, RPS6, RPS9 and 5.8S rRNA) or three (RPS2) independent experiments. **b**, PSMB2-FusionRed^{KI/KI} RPL29-eGFP^{AAVS1/AAVS1} cells were stimulated with 0.2 M sucrose for 30 min. Endogenous RPL29 was detected with anti-RPL29 antibody. Representative xy, xz and yz images of a single cell from two independent experiments. Scale bars, 5 μ m. **c**, PSMB2-FusionRed^{KI/KI} RPL29-eGFP^{AAVS1/AAVS1} cells were treated with the Pol I inhibitor CX-5461 (1 μ M, 5 h

prior), and stimulated with 0.2 M sucrose. Scale bars, 5 μ m. Right graphs indicate the foci numbers per cell and their diameters. RPL29 foci number are mean \pm s.d. from $n = 82$ cells (control) and $n = 68$ cells (CX-5461); **** $P < 0.0001$ by two-tailed unpaired t -test. RPL29 foci diameter are presented as mean \pm s.d. from $n = 542$ foci (control) and $n = 690$ foci (CX-5461); **** $P < 0.0001$ by two-tailed Mann-Whitney U -test. PSMB2 foci number are mean \pm s.d. from $n = 216$ cells (control) and $n = 228$ cells (CX-5461); ** $P = 0.0092$ by two-tailed Mann-Whitney U -test. PSMB2 foci diameter are mean \pm s.d. from $n = 815$ foci (control) and $n = 1,149$ foci (CX-5461); **** $P < 0.0001$ by two-tailed Mann-Whitney U -test. **d**, PSMB2-FusionRed^{KI/KI} cells transiently overexpressing (white arrows) or not overexpressing RPL7A-eGFP or RPS2-eGFP were stimulated with 0.2 M sucrose for 30 min. Scale bars, 5 μ m. Similar results were obtained from three independent experiments.



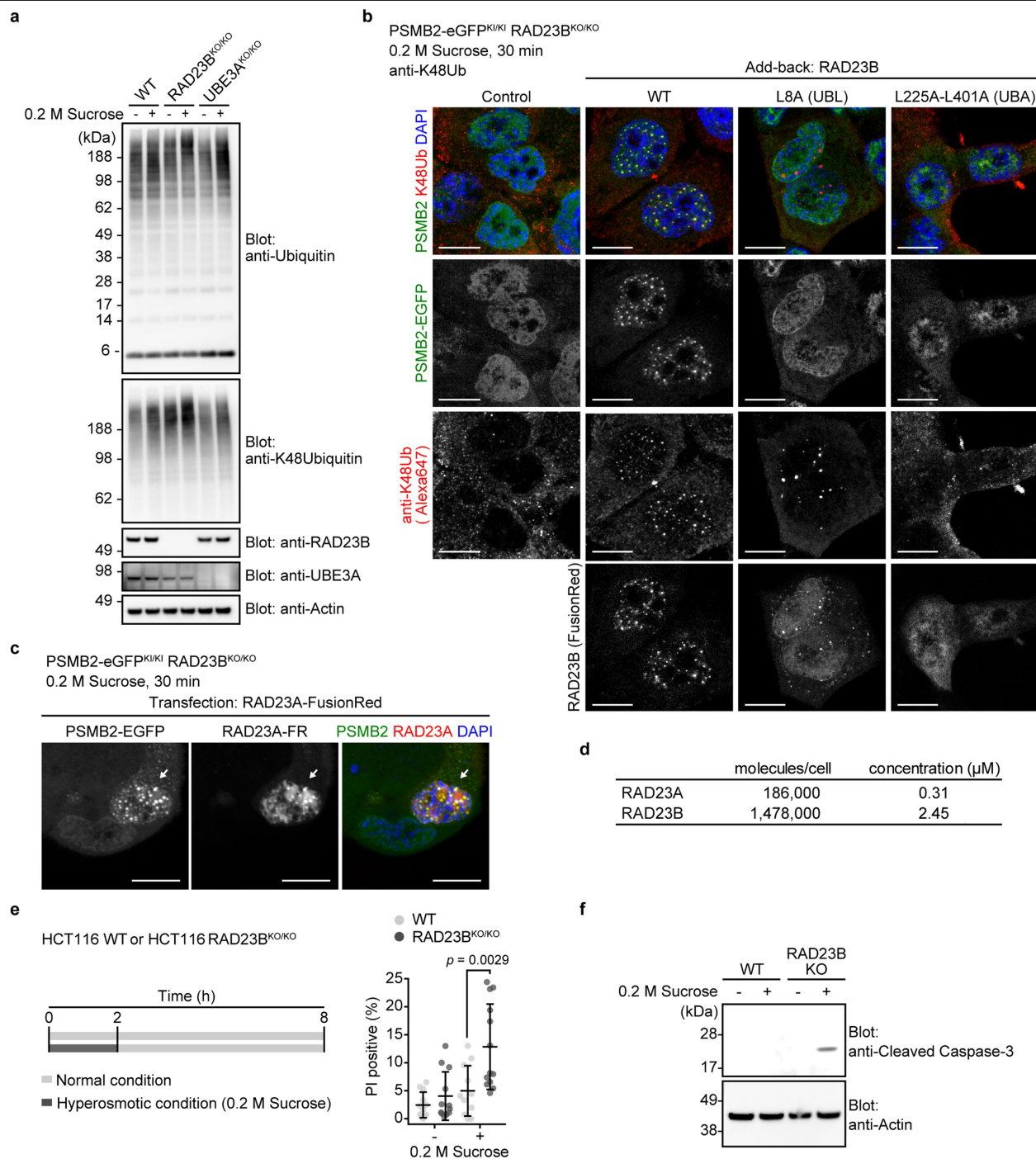
Extended Data Fig. 6 | Microscopy analysis of proteasome-interacting proteins. **a**, PSMB2-FusionRed^{KI/KI} cells were stimulated with 0.2 M sucrose for 30 min. Endogenous p97 and ubiquitin were detected with anti-p97 and anti-ubiquitin antibodies, respectively. Representative xy, xz and yz images of a single cell from two independent experiments. Scale bars, 5 μ m. **b**, PSMB2-FusionRed^{KI/KI} cells were stimulated with 0.2 M sucrose for 30 min. Endogenous p97 and RAD23B were detected with anti-p97 and anti-RAD23B antibodies, respectively. Representative xy, xz and yz images of a single cell from two

independent experiments. Scale bars, 5 μ m. **c**, PSMB2-eGFP^{KI/KI} cells were transiently transfected for 48 h with siRNA targeting *RAD23A*, *RAD23B*, *UBQLN1/2/4* (mixture) or *XPC*, or a control siRNA, and then stimulated with 0.2 M sucrose for 30 min. Right graph indicates the number of foci per cell (siControl, $n = 323$ cells; siRAD23A, $n = 129$ cells; siRAD23B, $n = 178$ cells; siUBQLN1/2/4, $n = 349$ cells; siXPC, $n = 339$ cells). Data are mean \pm s.d., **** $P < 0.0001$ by Kruskal–Wallis with Dunn's multiple comparisons test.



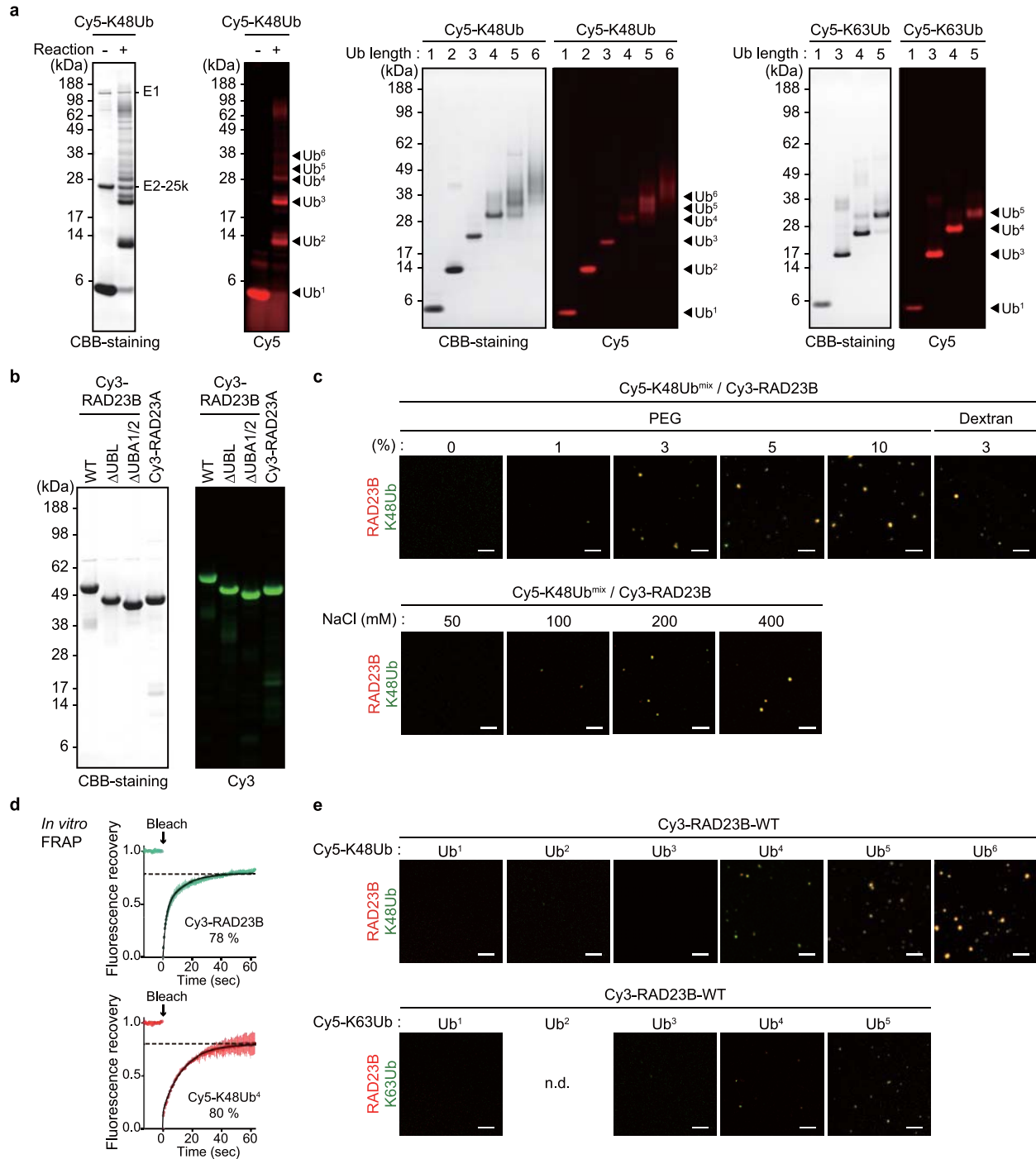
Extended Data Fig. 7 | Inhibition of proteasome or p97 activity results in enlargement of RPL29 condensates. a, PSMB2-FusionRed^{KI/KI}RPL29-eGFP^{AAVS1/AAVS1} cells treated with the proteasome inhibitor b-AP15 (5 μM, 10 min posterior), the p97 inhibitor NMS-873 (1 μM, 1 h prior), or the ubiquitin-activating enzyme (E1) inhibitor MLN-7243 (1 μM, 1 h prior), or lacking RAD23B, were stimulated with 0.2 M sucrose. Right graphs indicate the RPL29 foci numbers per cell and their diameters. The foci number per cell are presented as mean ± s.d. from $n=35$ cells (control), $n=44$ cells (b-AP15), $n=37$ cells (MLN-7243), $n=50$ cells (RAD23B KO) and $n=45$ cells (NMS-873). $P=0.2250$ (b-AP15), $**P=0.0085$ (MLN-7243), $**P=0.0017$ (RAD23B KO) and $P=0.9744$

(NMS-873) by one-way ANOVA with Dunnett's multiple comparisons test. The foci diameters are presented as mean ± s.d. from $n=220$ foci (control), $n=210$ foci (b-AP15), $n=132$ foci (MLN-7243), $n=170$ foci (RAD23B KO) and $n=271$ foci (NMS-873). $****P<0.0001$ (b-AP15), $****P<0.0001$ (MLN-7243), $****P<0.0001$ (RAD23B KO) and $****P<0.0001$ (NMS-873) by Kruskal–Wallis with Dunn's multiple comparisons test. Scale bars, 10 μm. **b**, Time-lapse images of single foci in the b-AP15 or NMS-873 treated PSMB2-FusionRed^{KI/KI}RPL29-eGFP^{AAVS1/AAVS1} cells under the same conditions described in **a**. Scale bars, 0.5 μm. Representative images from four (control) or two (b-AP15 and NMS-873) independent experiments.



Extended Data Fig. 8 | Characterization of RAD23B mutants in foci formation. **a**, Wild-type, RAD23B-KO (RAD23B^{KO/KO}) or UBE3A-KO (UBE3A^{KO/KO}) HCT116 cells were stimulated with 0.2 M sucrose for 30 min and immunoblotted with the indicated antibodies. Similar results were obtained from two independent experiments. **b**, PSMB2-eGFP^{KI/KI} RAD23B^{KO/KO} cells stably expressing FusionRed-fused RAD23B (WT), UBL (L8A) or UBA (L225A-L401A) were treated with 0.2 M sucrose for 30 min. Endogenous K48-linked ubiquitin chains were detected with K48-ubiquitin antibody and Alexa Fluor 647-labelled secondary antibody. Merged images represent K48-ubiquitin antibody (red), PSMB2-eGFP (green) and DAPI (blue). Scale bars, 10 μm. Representative images from two independent experiments. **c**, RAD23A-FusionRed-overexpressing PSMB2-eGFP^{KI/KI} RAD23B^{KO/KO} cells were treated with 0.2 M sucrose for 30 min. Scale bars, 10 μm. Similar results were obtained from two independent experiments. **d**, Cellular abundances of RAD23A and

RAD23B in HCT116 cells. See Source Data for details. **e**, Quantification of cell death by PI assay in wild-type or RAD23B-KO (RAD23B^{KO/KO}) HCT116 cells stimulated with 0.2 M sucrose for 2 h, followed by 6 h recovery under normal conditions. The graph indicates percentage of PI-positive cells. n represents the number of fields of view (untreated wild-type cells, $n = 12$, total 2,829 cells; sucrose-treated wild-type cells, $n = 11$, total 2,868 cells; untreated RAD23B^{KO/KO} cells, $n = 12$, total 2,865 cells; sucrose-treated RAD23B^{KO/KO} cells, $n = 13$, total 2,960 cells). Data are mean \pm s.d., $P = 0.0029$ by one-way ANOVA with Tukey's multiple comparison test. **f**, Western blot analysis of wild-type or RAD23B-KO HCT116 cells under the same conditions described in **e**. The presence of cleaved caspase-3 was used as a marker for apoptosis. Representative results from two independent experiments. Gel source data for **a**, **f** are shown in Supplementary Fig. 1.



Extended Data Fig. 9 | Liquid-liquid phase separation of ubiquitin chains and RAD23B in vitro. **a**, SDS-PAGE and fluorescent images of Cy5-labelled proteins used in the in vitro liquid-liquid phase separation assay. Cy5-labelled K48-linked ubiquitin chains (Cy5-K48Ub^{mix}) generated by enzymatic reactions using E2-25K (left), Cy5-labelled K48-linked ubiquitin chains size-fractionated by gel filtration (middle) and Cy5-labelled K63-linked ubiquitin chains size-fractionated by gel filtration (right). Note that K63-Ub2 was omitted owing to difficulty of separation. **b**, SDS-PAGE and fluorescent images of Cy3-labelled RAD23 proteins. **c**, Effects of concentration of molecular crowding agents and NaCl. Images of the solution were obtained 90 min after mixing

20 μM Cy5-K48Ub^{mix} and 20 μM Cy3-RAD23B(WT) with the indicated conditions. Scale bars, 5 μm. **d**, Average of quantification of fluorescence recovery of 20 μM Cy3-RAD23B and 60 μM Cy5-K48Ub⁴ (5 min after mixing in 10% PEG) and double-exponential fitting curve with s.e.m. ($n = 2$). **e**, Chain length-dependent formation of liquid droplets. Images of the solution were obtained 90 min after mixing in 20 μM size-fractionated Cy5-K48Ub, Cy5-K63Ub, or ubiquitin monomer (Cy5-Ub) and 20 μM Cy3-RAD23B(WT). All samples contained 3% PEG in 200 mM NaCl. Scale bars, 5 μm. Representative results from two independent experiments. Gel source data for **a**, **b**, are shown in Supplementary Fig. 1.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The images of the live cells (Fig. 1a, 2d, 3a; Extended Data Fig. 2b, 7b; Supplementary Video 1, 2, 3, 4, 5, 6) and the two-color or three-color samples of the fixed cells (Fig. 1a, 1c, 2c, 3b, 4a, 4c; Extended Data Fig. 1a, 2a, 2c, 2e, 3a, 3b, 3c, 5a, 5c, 5d, 6c, 7a, 8c) were captured with iQ2.9.1 (Andor) or Metamorph 7.8 (Molecular Devices). The FRAP images (Fig. 3c) and four-color samples of the fixed cells (Extended Data Fig. 5b, 6a, 6b, 8b) were captured with LAS X 2.0.1 (Leica). The time-lapse images (Fig. 1d) were captured with CV1000 Software 1.06.06 (YOKOGAWA). Cryo-EM images (Fig. 1b; Extended Data Fig. 2d) were collected using SerialEM (V3.7). TEM images (Fig. 2a; Extended Data Fig. 4c, 4d) were collected using AnalySIS 5.0 (Olympus SIS). Droplet formation (Fig. 4d, 4e, 4f; Extended Data Fig. 9c, 9d, 9e) were captured with FV31S-SW 2.3.1.163 (Olympus). Fluorescence images from CLEM (Extended Data Fig. 4b) were collected using iQ2.9.1 (Andor). MS data were acquired by Xcalibur software 2.2 (Thermo Fisher Scientific).

Data analysis

All image analysis and quantification were performed with Metamorph 7.8 (Molecular Devices). All statistical analysis were performed with GraphPad Prism7 (GraphPad Software). Cryo-EM images were processed using MATLAB R2015b (Mathworks) and TOM software toolbox (V6.1). K2 frames were aligned using in-house software (K2Align). Tomograms were reconstructed from aligned tilt series by using the IMOD (4.9.0) software package. PyTom (0.97) was used to identify proteasomes in the tomograms. The resulting subtomograms were cropped out, CTF- corrected and classified using RELION (3.0). Fluorescence images and TEM images for CLEM were merged using Icy 1.9.7.0 (Institut Pasteur). MS data were analyzed using Mascot search program (Matrix Science) in Proteome Discoverer 1.3 (Thermo Fisher Scientific).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that the data supporting the findings of this study are available within the paper. Supplementary data (Supplementary Video1, 2, 3, 4, 5, 6, 7; and Supplementary Figure 1) and Source data for all quantification (Fig. 1c, 1d, 2c, 3a, 3b, 3c, 4a, 4b, 4d; Extended Data Fig. 1e, 3a, 3b, 3c, 5a, 5c, 6c, 7a, 8d, 8e, 9d) are provided with the online version of the paper. Supplementary Figure 1 contains uncropped images of blots and gels. Raw images of Cryo-ET have been deposited to the Electron Microscopy Data Bank under accession code EMD-10494. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD01637 and PXD016369. All other datasets analysed in the current study are available from the corresponding author upon reasonable requests.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not compute statistical analyses to predetermine sample sizes. Sample sizes were chosen based on similar published studies elsewhere. All analyzes were at least two and more independent samples (for n, see figure legends), and in each experimental repeat yielding highly similar results.
Data exclusions	No data exclusions.
Replication	We have reproduced all our results in a minimum of two independent experimental repeats. Experimental findings have been either reproduced independently by co-authors of the paper using the same experimental conditions.
Randomization	Randomization was not required as no human participants or animal models were reported in this manuscript and experiments were performed on specified cell lines.
Blinding	Many microscopy images were blinded and analyzed in an automated fashion.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Antibody (clone), dilution, Supplier, Cat. number, Lot number:
 Ubiquitin, 1:500 (for IF) or 1:1000 (for WB), Dako, Z0458, 00082271
 Ubiquitin (FK2), 1: 500, Nippon Bio-Test Laboratories, NBT-MFK003, C7070
 Ubiquitin Lys48-Specific (Apu2), 1:500 (for IF) or 1:1000 (for WB), Millipore, 05-1307, 2918991
 Ubiquitin Lys63-Specific (Apu3), 1:500, Millipore, 05-1308, 3137755
 PSMB2, 1:1000, Enzo, BML-PW8890, Z03617b

PSMD6, 1:1000, Enzo, BML-PW8225-0100, 01041704
 PSMA1, 1:500, Kumatori A, Tanaka K et al. PNAS, 1990
 VCP (5), 1:500, Abcam, ab11433, GR3180147-3
 RAD23B (H-120), 1:500, Santa Cruz, sc-67225, A0809
 RAD23B (D4W7F), 1:1000, Cell Signaling Technology, 13525, 1
 RAD23A (D7U7Z), 1:1000, Cell Signaling Technology, 24555, 1
 UBE3A, 1:100 (for IF) or 1:1000 (for WB), Protein Tech, 10344-1-AP, 00010503
 RPL15, 1:500, Protein Tech, 16740-1-AP, 00008177
 RPL7A (E109), 1:100, Cell Signaling Technology, 2415, 2
 RPL29 (B01P), 1:50, Abnova, H00006159-B01P), GB251
 RPL35, 1:50, Sigma, SAB4500233, 3114173
 RPL4 (RQ-7), 1:50, Santa Cruz, sc-100838, A3017
 RPS2 (EPR10834(B)), 1:100, Abcam, ab155961, GR117602-1
 RPS9, 1:100, Proteintech, 18215-1-AP, 00009939
 RPS6 (5G10), 1:200, Cell Signaling Technology, 2217, 7
 rRNA (Y10b), 1:100, Santa Cruz, sc-33678, A0510
 PML (PG-M3), 1:100, Santa Cruz, sc-966, A0313
 Coilin, 1:1000, Abcam, ab11822, GR127265-2
 BMI1 (D20B7), 1:600, Cell Signaling Technology, 6964, 1
 SC35, 1:100, Abcam, ab11826, GR144730-2
 CENPC, 1:1000, MBL, PD030, 004
 phospho-Histone H2A.X (Ser139) (JBW301), 1:100, Millipore, 05-636, 2310335
 Cleaved Caspase-3 (Asp175) (5A1E), 1: 1000, Cell Signaling Technology, 9661, 42
 FLAG (M2), 1:2000, Sigma, A8592, SLBH1183V
 Actin, 1:4000, MBL, PM053-7, 007
 Alexa Fluor 488–conjugated anti–mouse, 1:1000, Invitrogen, A-11029, 1252783
 Alexa Fluor 488–conjugated anti–rabbit, 1:1000, Invitrogen, A-11034, 1073084
 Alexa Fluor 568–conjugated anti–mouse, 1:1000, Invitrogen, A-11031, 1736975
 Alexa Fluor 568–conjugated anti–rabbit, 1:1000, Invitrogen, A-11036, 1832035
 Alexa Fluor 568–conjugated anti–guinea pig, 1:1000, Invitrogen, A-11075, 1170591
 Alexa Fluor 647–conjugated anti–mouse, 1:1000, Invitrogen, A-21236, 1511347
 Alexa Fluor 647–conjugated anti–rabbit, 1:1000, Invitrogen, A-21245, 1892148A
 HRP-conjugated goat anti-mouse Ig, 1:10000, Jackson ImmunoResearch Laboratories, 115-035-003
 HRP-conjugated goat anti-rabbit Ig, 1:10000, Jackson ImmunoResearch Laboratories, 111-035-144, 55285

Validation

Ubiquitin (Dako, Z0458): Previously used for publication (Tsuchiya H., Mol Cell. 2017 May 18;66(4):488-502.e7.), and (Kageyama M., PLoS One. 2019 May 31;14(5):e0217945.).
 Ubiquitin (FK2, Nippon Bio-Test Laboratories, 0918-2): Previously used for publication (Nozawa T et al., Autophagy. 2017;13(11):1841-1854.).
 Ubiquitin, Lys48-Specific (Millipore, 05-1307): Supplier has been validated in Flow cytometry analysis of Jurkat cells, Immunocytochemistry of NIH/3T3, HeLa, and A431 cells, Immunoprecipitation via precipitated ubiquitin-containing proteins from HeLa cell lysates, Immunohistochemistry of rat brain sections and human breast carcinoma, and Western blotting of the inhibition assay using HeLa cell lysate.
 Ubiquitin, Lys63-Specific (Millipore, 05-1308): Supplier has been validated in Flow cytometry analysis of Jurkat cells, Immunocytochemistry of staining of colorectal cancer tissue sections limited to tumor cells, Immunoprecipitation via precipitated ubiquitin-containing proteins from HeLa cell lysates, Immunohistochemistry of staining of colorectal cancer tissue sections limited to tumor cells, and Western blotting of the inhibition assay using HeLa cell lysate.
 PSMB2 (Enzo, BML-PW8890-0100): Previously used for publication (Zana Lukic et al., Retrovirology. 2011; 8: 93.)
 PSMD6 (Enzo, BML-PW8225-0100): Previously used for publication (Mata-Cantero L et al., J Proteomics. 2016 Apr 29;139:45-59.)
 PSMA1: Previously used for publication (Kumatori A, Tanaka K et al., Proc Natl Acad Sci U S A. 1990 Sep;87(18):7071-5.)
 VCP (Abcam, ab11433): Supplier has been validated in Western blotting of cell lysate from mouse embryonic fibroblasts and CA46 cells, Immunohistochemistry of Human colon carcinoma tissues limited to tumor cells, Flow cytometry analysis of Platelets, and Immunocytochemistry of WiDr colon carcinoma cells.
 RAD23B (Santa Cruz, sc-67225): Supplier has been validated in Immunohistochemistry of human cerebellum, and Immunocytochemistry of human cell line U-2 OS cells.
 RAD23B (Cell Signaling Technology, 13525): Supplier has been validated in Western blotting of extracts from 293T cells.
 RAD23A (Cell Signaling Technology, 24555): Supplier has been validated in Western blotting of extracts from SK-OV-3, OVCAR3, CCRF-CEM, RPMI 8226, 786-O, A498, Hepa 1-6, M-1, C6, A-10, COS-7 and 293T cells.
 UBE3A (Protein Tech, 10344-1-AP): Supplier has been validated in Western blotting of mouse brain tissue or extracts from HEK293, HeLa, Jurkat and K562 cells, Immunohistochemistry of paraffin-embedded human lung cancer tissue, and Immunocytochemistry of HeLa cells.
 RPL15 (Protein Tech, 16740-1-AP): Supplier has been validated in Western blotting of COLO 320 cells, and Immunocytochemistry of HeLa cells.
 RPL7a (Cell Signaling Technology, 2415): Supplier has been validated in Western blotting of extracts from HEK293, NIH/3T3, PC12 and COS cells, and Immunocytochemistry of COS cells.
 RPL29 (Abnova, H00006159-B01P): Supplier has been validated in Western blotting of extracts from human pancreas, Immunocytochemistry of HeLa cells.
 RPL35 (Sigma, SAB4500233): Supplier has been validated in Immunoblotting analysis of extracts from 293, HeLa, COLO and Jurkat cells.
 RPL4 (Santa Cruz, sc-100838): Supplier has been validated in Immunoblotting analysis of extracts from HeLa, Jurkat and K-562 cells, and Immunocytochemistry of HeLa cells.
 RPS2 (Abcam, ab155961): Supplier has been validated in Immunohistochemistry of Human kidney and pancreas tissue, Western blotting of the inhibition assay using HeLa, A549 and 293T cell lysate, and Immunocytochemistry of HeLa cells.
 RPS9 (Proteintech, 18215-1-AP): Supplier has been validated in Western blotting of extracts from mouse uterus tissue, Immunocytochemistry of Human kidney, and Immunocytochemistry of HeLa cells.

RPS6 (Cell Signaling Technology, 2217): Supplier has been validated in Western blotting of extracts from HeLa, NIH/3T3, PC12 and COS cells, Immunocytochemistry of human carcinoma, and Immunocytochemistry of mouse brain.

rRNA (Santa Cruz, sc-33678): Supplier has been validated in Immunocytochemistry of HeLa cells.

PML (Santa Cruz, sc-966): Supplier has been validated in Western blotting of extracts from K-562 and COLO cells, Immunohistochemistry of human urinary bladder tissue, and Immunocytochemistry of A-431 cells.

Coilin (Abcam, ab11822): Supplier has been validated in Immunocytochemistry of MCF-7 cells.

BM11 (Cell Signaling Technology, 6964): Supplier has been validated in Western blotting of extracts from HeLa, A-549, COS-7 and Vero cells, Immunocytochemistry of human lung carcinoma, and Immunocytochemistry of COS-7 cells.

SC35 (Abcam, ab11826): Supplier has been validated in Immunocytochemistry of HEK-293 human kidney cells, and Immunocytochemistry of human hippocampus tissue.

CENPC (MBL, PD030): Supplier has been validated in Western blotting of extracts from HeLa cells, and Immunocytochemistry of HeLa cells.

Phospho-Histone H2A.X (Ser139) (Millipore, 05-636): Supplier has been validated in Western blotting of extracts from Jurkat cells, Immunocytochemistry of Jurkat cells, and Immunohistochemistry of RNF168-WT and RNF 168-SA/SEKI mice lung tissue. Previously used for publication (Xe, X., et al. (2015) Nat. Cell Biol. 20 (3); 320-331.), (Meier, Andreas, et al. EMBO J., 26: 2707-18. (2007))

Cleaved Caspase-3 (Asp175) (Cell Signaling Technology, 9661): Supplier has been validated in Western blotting of extracts from HeLa, NIH/3T3 and C6 cells, Immunohistochemistry of human tonsil and mouse embryo, Immunocytochemistry of HT-29 cells, and Flow cytometry analysis of Jurkat cells.

FLAG (Sigma, A8592): Previously used for publication (Tsuchiya H et al., Nat Commun. 2018 Feb 6;9(1):524.), and (Chaumet A et al., Nat Commun. 2015 Sep 10;6:8218.).

Actin (MBL, PM053-7): Supplier has been validated in Western blotting of extracts from PC12 and HEK293 cells.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HCT116, hTERT RPE-1, HEK293T, and E14TG2a cells were purchased from American Type Culture Collection (ATCC).
Authentication	Cell line authentication was not performed.
Mycoplasma contamination	All cell lines used were periodically checked by fluorescent microscopy. Cells were stained with DAPI and examined under X60 or X100 objective lens and no contamination was found.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Phase separation organizes the site of autophagosome formation

<https://doi.org/10.1038/s41586-020-1977-6>

Received: 17 December 2018

Accepted: 2 December 2019

Published online: 5 February 2020

Yuko Fujioka¹, Jahangir Md. Alam¹, Daisuke Noshiro^{1,2}, Kazunari Mouri³, Toshio Ando², Yasushi Okada^{3,4}, Alexander I. May^{5,6}, Roland L. Knorr^{7,8,9}, Kuninori Suzuki^{10,11}, Yoshinori Ohsumi⁵ & Nobuo N. Noda^{1*}

Many biomolecules undergo liquid–liquid phase separation to form liquid-like condensates that mediate diverse cellular functions^{1,2}. Autophagy is able to degrade such condensates using autophagosomes—double-membrane structures that are synthesized de novo at the pre-autophagosomal structure (PAS) in yeast^{3–5}. Whereas Atg proteins that associate with the PAS have been characterized, the physicochemical and functional properties of the PAS remain unclear owing to its small size and fragility. Here we show that the PAS is in fact a liquid-like condensate of Atg proteins. The autophagy-initiating Atg1 complex undergoes phase separation to form liquid droplets in vitro, and point mutations or phosphorylation that inhibit phase separation impair PAS formation in vivo. In vitro experiments show that Atg1-complex droplets can be tethered to membranes via specific protein–protein interactions, explaining the vacuolar membrane localization of the PAS in vivo. We propose that phase separation has a critical, active role in autophagy, whereby it organizes the autophagy machinery at the PAS.

The PAS is a transient structure that is regulated by nutrient conditions and invariably forms on the vacuole in yeast on starvation³. The PAS initially comprises Atg1 complexes consisting of Atg1, Atg13, Atg17, Atg29 and Atg31, which are abundant with intrinsically disordered regions (IDRs)⁶. This ‘early PAS’ then matures by recruiting downstream Atg proteins and vesicles, subsequently serving as the site of autophagosome formation^{7,8}. These features are consistent with biomolecular condensates (also known as membraneless organelles) that are formed through liquid–liquid phase separation^{1,2}. We therefore set out to determine whether the PAS is in fact a biomolecular condensate, and examined whether condensate formation is able to explain its spatiotemporal behaviour in the cell.

The PAS is a liquid-like condensate

First, we studied the dynamics of the PAS using fluorescence microscopy. We used yeast cells overexpressing GFP–Atg13 in an *atg11Δ* background to ensure that fluorescence intensity is sufficient for quantitative analysis and that the PAS is formed in response to starvation rather than the Atg11-dependent pathway, which is constitutive and responsible for cytoplasm-to-vacuole targeting^{9,10}. We confirmed that overexpression of GFP–Atg13 did not impair autophagy activity (Extended Data Fig. 1a). Upon nitrogen starvation, GFP–Atg13 formed puncta that dissolved within 8 min of addition of a nitrogen source (Extended Data Fig. 1b), consistent with a previous study⁹, suggesting that the PAS is a dynamic and transient entity. Fluorescence recovery

after photobleaching (FRAP) experiments showed a quick recovery of fluorescence in puncta (corresponding to the PAS) containing GFP–Atg13, with a recovery half-time of 1.3 s after photobleaching (Fig. 1a, Extended Data Fig. 1c, Supplementary Video 1). This exchange rate is comparable with or even faster than that of molecules in nuclear biomolecular condensates¹¹. We also observed rapid fluorescence recovery of Atg1–GFP, Atg13–GFP and Atg17–GFP proteins expressed at endogenous levels (Extended Data Fig. 1d). We next performed fluorescence-correlation microscopy (FCS) experiments on GFP–Atg13. The diffusion coefficient of GFP–Atg13 in the PAS was about half that in the cytosol (Fig. 1b, c, Extended Data Fig. 1e). These results provide evidence supporting a dynamic liquid-like structure of the PAS in which GFP–Atg13 can move diffusively, although the crowded environment of the PAS would have slowed down the diffusion, as reported with other liquid-droplet structures^{12,13}. Owing to the small size of the PAS, the rate determined by FCS could potentially include both entry and exit of GFP–Atg13 molecules to the PAS. We next generated and performed FRAP analyses on a giant PAS (diameter >1 μm) by overexpressing GFP–Atg13 from a multicopy plasmid; Atg13–GFP fluorescence rapidly recovered (less than 1 s) in a partially quenched region (Fig. 1d, Extended Data Fig. 1f, Supplementary Video 2). These data show that Atg13 is not only able to enter and exit the PAS, but also moves freely within the PAS.

Previous studies have established that 1,6-hexanediol is able to inhibit liquid–liquid phase separation of biomolecules¹⁴. When yeast cells were treated with 1,6-hexanediol, Atg13 puncta rapidly dissolved; they reappeared on the vacuolar membrane after removal of 1,6-hexanediol

¹Institute of Microbial Chemistry (BIKAKEN), Tokyo, Japan. ²Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kanazawa, Japan. ³Center for Biosystems Dynamics Research (BDR), RIKEN, Osaka, Japan. ⁴Department of Physics, Universal Biology Institute (UBI) and International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, Tokyo, Japan. ⁵Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan. ⁶Tokyo Tech World Research Hub Initiative (WRHI), Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan. ⁷Department of Theory and Bio-Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany. ⁸Graduate School and Faculty of Medicine, The University of Tokyo, Tokyo, Japan. ⁹Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany. ¹⁰Life Science Data Research Center, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan. ¹¹Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo, Japan. *e-mail: nn@bikaken.or.jp

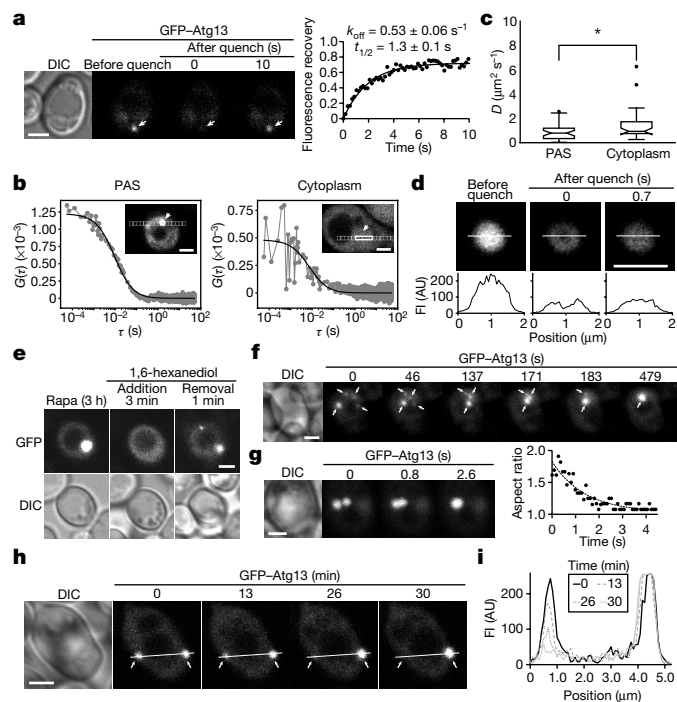


Fig. 1 | The PAS behaves as a liquid droplet in vivo. **a**, Left, rapid recovery of fluorescence of GFP-Atg13 puncta after photobleaching. Right, fluorescence recovery fitted to a curve; dissociation rate constant (k_{off}) and recovery half-time ($t_{1/2}$) values (mean \pm s.d.) were calculated from $n = 3$ independent experiments. DIC, differential interference contrast microscopy. **b**, FCS measurements of GFP-Atg13 in the PAS and in the cytoplasm. The autocorrelation function of the fluorescence signal is calculated for a single pixel near the centre of the PAS, or for the five pixels in the cytoplasm (insets). The autocorrelogram was fitted to a three-dimensional one-component diffusion model to estimate diffusion coefficients. **c**, Tukey-style box-and-whisker plot²⁸ of the diffusion constants of GFP-Atg13 in the PAS ($n = 35$ cells) and in the cytoplasm ($n = 32$ cells) measured by scanning FCS (see Methods for details). The 95% confidence interval of the medians are shown with notches, and are estimated to be $0.79 \pm 0.23 \mu\text{m}^2 \text{s}^{-1}$ (PAS) and $0.93 \pm 0.27 \mu\text{m}^2 \text{s}^{-1}$ (cytoplasm). $*P = 0.031$, two-sided Wilcoxon Mann-Whitney U -test. **d**, Partial FRAP experiment with a giant PAS. Graphs indicate line profiles of fluorescence intensity (FI) in the above images. **e**, Reversible effect of 1,6-hexanediol on the formation of Atg13-GFP puncta. Rapa, rapamycin treatment. **f**, Appearance of multiple Atg13-GFP puncta and their coalescence to form one large punctum. **g**, Left, coalescence of two PAS precursors. Right, graph shows the change in aspect ratio during coalescence. **h**, Ostwald ripening of PAS precursors observed in vivo. **i**, Line profile of fluorescence intensity in **h**. All scale bars are $2 \mu\text{m}$. Experiments were repeated independently twice (**e**, **f**) or three times (**a**, **d**, **g**, **h**) with similar results.

(Fig. 1e) and coalesced to form a larger punctum (Extended Data Fig. 1g, Supplementary Video 3). To observe the process of PAS formation in detail, we optimized the expression level of GFP-Atg13 using the inducible *GAL1* promoter. When yeast cells accumulating GFP-Atg13 during a 7-h induction were treated with rapamycin for 10 min, multiple small GFP-Atg13 puncta appeared and coalesced to form a large punctum (Fig. 1f, Supplementary Video 4). Upon coalescence of two puncta, the aspect ratio changed from approximately 2.0 to 1.0 within a few seconds, reflecting the liquid-like nature of puncta (Fig. 1g, Extended Data Fig. 1h, Supplementary Video 5). We occasionally observed the enlargement of one PAS punctum coinciding with the reduction of another (Fig. 1h, i, Extended Data Fig. 1i, Supplementary Video 6). This phenomenon is consistent with Ostwald ripening, although local phosphorylation events might dissolve the shrinking PAS punctum. Collectively, these data suggest that the starvation-induced PAS is a liquid-like biomolecular condensate that is formed by liquid-liquid phase separation.

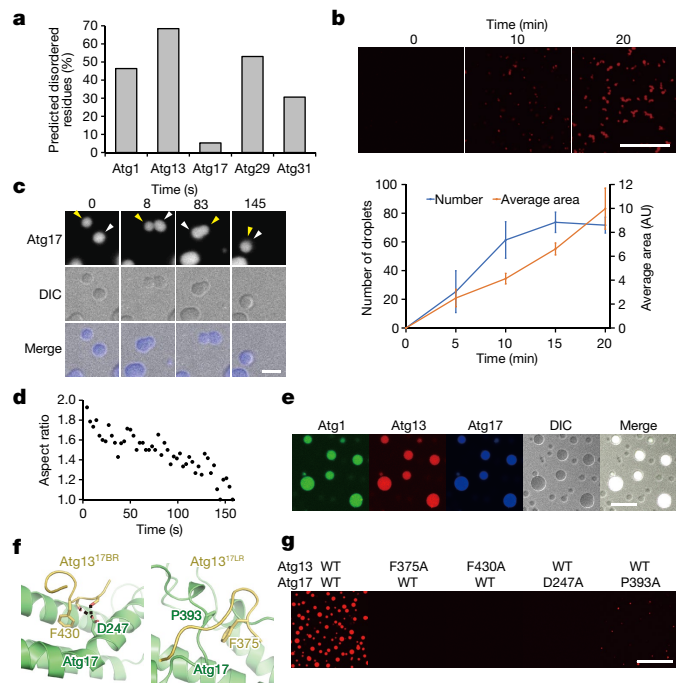


Fig. 2 | The Atg1 complex undergoes phase separation in vitro. **a**, Percentage of residues predicted to be disordered by DISOPRED²⁹. **b**, Formation of liquid droplets of the Atg1 complex on mixing. The bottom graph shows the mean \pm s.d. of number and area of the droplets ($n = 3$ independent experiments). Scale bar, $50 \mu\text{m}$. **c**, Coalescence observed between the droplets of the Atg1 complex. Scale bar, $5 \mu\text{m}$. **d**, Time course of the changes in aspect ratio during coalescence in **c**. **e**, Co-occurrence of Atg1, Atg13 and Atg17 in the Atg1-complex droplets. Scale bar, $30 \mu\text{m}$. **f**, Interactions of Atg13 17BR (left) and 17LR (right) with Atg17 (PDB 5JHF). Residue numbers refer to the *Saccharomyces cerevisiae* Atg17 sequence. **g**, Formation of scaffold-complex liquid droplets and their inhibition by mutations. Scale bar, $50 \mu\text{m}$. Experiments were repeated independently three times with similar results (**c**, **e**, **g**).

The Atg1 complex forms droplets in vitro

We previously reported that a higher-order assemblage of the Atg1 complex organizes the PAS and initiates autophagy⁶. With the exception of Atg17, the components of the Atg1 complex contain many IDRs (Fig. 2a, Extended Data Fig. 2a). We speculated that the higher-order assemblage of the Atg1 complex induces liquid-liquid phase separation, providing a mechanism for formation of a liquid-like PAS. SNAP-tagged Atg1, Atg13, and Atg17-Atg29-Atg31 were purified (Extended Data Fig. 2b) and labelled with distinct fluorescent dyes. Mixing of these proteins resulted in the immediate onset of phase separation, indicated by the appearance of multiple spherical droplets, the number of which reached a maximum at 15 min before a subsequent decline, whereas the total area occupied by the droplets increased continuously (Fig. 2b). Again, droplets coalesced to form a larger spherical droplet, during which the aspect ratio changed from approximately 2.0 to 1.0, suggesting a liquid-like state (Fig. 2c, d, Extended Data Fig. 2c). Droplets emitted three distinct fluorescence signals corresponding to Atg1, Atg13 and Atg17-Atg29-Atg31, indicating that these components colocalize within droplets (Fig. 2e). Components of biomolecular condensates can be divided into two qualitative classes: scaffolds, which are essential for the formation of condensates, and clients, which are dispensable⁴. As Atg1 was dispensable for droplet formation (Extended Data Fig. 2d, buffer), Atg1 can be considered as a client, whereas the other components act as scaffolds; we therefore refer to the Atg13-Atg17-Atg29-Atg31 droplets as scaffold droplets. These scaffold droplets were promptly dispersed by 1,6-hexanediol treatment (Extended Data Fig. 2d, e), similar to the PAS in vivo (Fig. 1e). Phase separation was

most efficient at pH 6.0 and was impaired under higher pH conditions (>7.0) (Extended Data Fig. 2f, g), consistent with PAS formation under starvation conditions that result in acidification of the cytoplasm of budding yeast¹⁵ to a pH around 6.0.

It is known that liquid–liquid phase separation of proteins is induced by two distinct mechanisms: the first involves nonspecific weak interactions between IDRs, whereas the second is brought about by multiple specific interactions between proteins that possess numerous binding modules². Atg13 possesses both an Atg17-binding region (17BR) and an Atg17-linking region (17LR), which bind to distinct regions in Atg17 (Fig. 2f). Formation of scaffold droplets was severely impaired by the F430A or F375A mutation in Atg13 and the D247A or P393A mutation in Atg17 (Fig. 2g, Extended Data Fig. 2h), which attenuate the Atg17–17BR and Atg17–17LR interactions that are essential for PAS formation and autophagy^{6,16}. Thus, phase separation of the Atg1 complex is facilitated by multiple specific interactions between Atg13 and Atg17, both in vitro and in vivo. The 1:1 stoichiometry of Atg13 and Atg17–Atg29–Atg31 is optimal for phase separation, which is impaired by excess Atg17–Atg29–Atg31 (Extended Data Fig. 2i). In line with this, overexpression of Atg17–GFP impaired droplet liquidity (Extended Data Fig. 2j). Collectively, these data suggest that the Atg1 complex undergoes phase separation to form a liquid droplet through two-site binding between Atg13 and Atg17, an essential mechanism of PAS formation in vivo⁶.

Regulation of phase separation

Under nutrient-rich conditions, Atg13 is highly phosphorylated by TORC1, which inhibits the formation of the Atg1 complex and the PAS^{16,17}. Some Atg proteins accumulate en masse at the PAS when the kinase activity of Atg1 is inhibited¹⁰. These observations suggest that phosphorylation events negatively regulate the formation of the PAS. When Atg13 was incubated with TORC1 purified from yeast (Extended Data Fig. 3a) and ATP, Atg13 was hyperphosphorylated, including on Ser428 and Ser429, whose phosphorylation impairs the interaction of Atg13 with Atg17 as well as PAS formation¹⁶ (Fig. 3a). Phosphorylated Atg13 lost the ability to form droplets with Atg17–Atg29–Atg31 (Fig. 3b), indicating that TORC1 inhibits phase separation by directly phosphorylating Atg13, especially at Ser428 and Ser429.

The kinase activity of Atg1, which is essential for autophagy progression¹⁷, is activated upon starvation. Activation of Atg1 requires autophosphorylation of the kinase domain at Thr226, which is markedly enhanced following rapamycin treatment¹⁸. Clustering of Atg1 by a selective autophagy cargo or by its targeting to the vacuole accelerates the autophosphorylation^{19,20}. We monitored phosphorylation of Thr226 using a phosphorylation-specific antibody for Thr226¹⁸ when Atg1 alone, phase-separated Atg1 complex or an Atg1 complex in which phase separation was inhibited by an F430A mutation in Atg13 were incubated with ATP (Extended Data Fig. 3b). There was no increase in autophosphorylated protein for Atg1 alone, but a mild increase was observed in the Atg1(F430A) complex, and a more marked increase was observed in the phase-separated Atg1 complex (Fig. 3c, Extended Data Fig. 3c). These data suggest that phase separation of the Atg1 complex facilitates activation of Atg1 kinase, possibly by increasing collisions between individual Atg1 molecules.

Next, we studied the effect of Atg1 kinase activity on droplet formation. Addition of ATP induced phosphorylation of Atg1, Atg13 and Atg29 within 10 min in the Atg1 complex, but not in the Atg1(D211A) kinase-dead complex, indicating that these three proteins were phosphorylated by Atg1 (Extended Data Fig. 3d, e). Incubation with ATP dissolved the droplets of the wild-type Atg1 complex, but not those of the Atg1(D211A) complex, in a similar time frame (Extended Data Fig. 3f), confirming that Atg1-mediated phosphorylation, but not the activity of ATP as a hydrotrope²¹, inhibited phase separation of the Atg1 complex.

In contrast to the in vitro observation that Atg1-complex droplets dissolve within minutes in the presence of Atg1 kinase activity, the

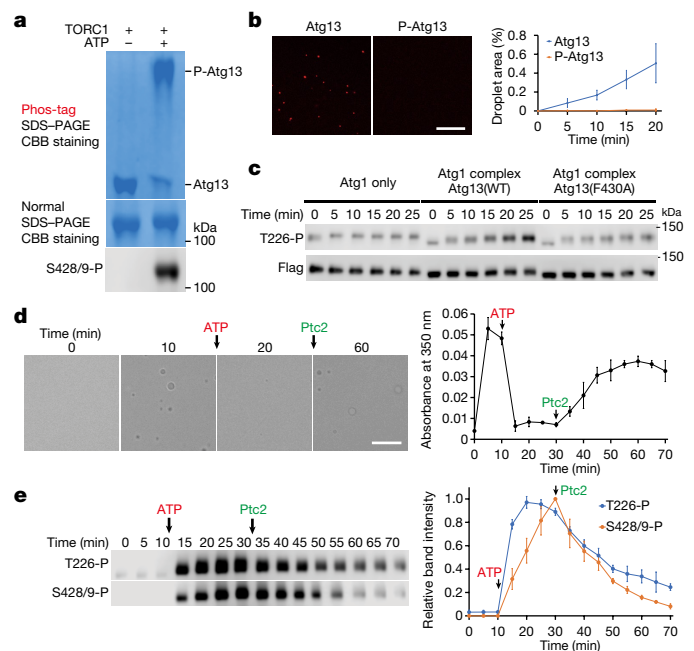


Fig. 3 | Phase separation of the Atg1 complex is dynamically controlled by phosphorylation-dependent regulation of Atg13. **a**, Phosphorylation of Atg13 (P-Atg13) by TORC1. CBB, Coomassie brilliant blue (CBB) staining. **b**, Left, impairment of phosphorylated Atg13 phase separation on mixing with Atg17–Atg29–Atg31. Scale bar, 30 μ m. Right, graph shows time-course analysis of droplet area. Data are mean \pm s.d. ($n = 3$ independent experiments) (**b**, **d**, **e**). **c**, Enhancement of Atg1 Thr226 phosphorylation on phase separation in vitro. **d**, Left, effect of ATP and Ptc2 on phase separation of the Atg1 complex. Right, graph shows absorbance at 350 nm as an indicator of droplet formation. Scale bar, 10 μ m. **e**, Left, phosphorylation of Atg1 at Thr226 and Atg13 at Ser428/429 as assessed by western blot analyses. Right, relative band intensity quantified from blots. Experiments were repeated independently three times with similar results (**a**, **c**). For gel source data, see Supplementary Fig. 1 (**c**, **e**).

PAS continues to exist for several hours despite its activation of Atg1 kinase. The PP2C phosphatases Ptc2 and Ptc3 have been reported to promote PAS formation and autophagy by dephosphorylating Atg1 and Atg13²². This suggests that the balance of phosphorylation and dephosphorylation of Atg13 may be important for the maintenance of the PAS. We studied the effect of Ptc2 on phase separation of Atg1 complex (Fig. 3d). Addition of ATP promptly dissolved Atg1-complex droplets. Further addition of recombinant Ptc2 (Extended Data Fig. 3g, h) gradually regenerated the droplets, indicating that Ptc2-mediated dephosphorylation promotes phase separation of the Atg1 complex. Upon ATP addition, phosphorylation of Atg1 at residue Thr226, and shortly afterwards, of Atg13 at Ser428 and Ser429 (Ser428/429) occurred. Thr226 and Ser428/429 were then dephosphorylated after addition of Ptc2 (Fig. 3e). As Ser428/429 of Atg13 is the most critical phosphorylation site for inhibiting PAS formation¹⁶, these results suggest that Atg1 and Ptc2 regulate the reversible phase separation of the Atg1 complex through phosphorylating and dephosphorylating Atg13 Ser428/429. Notably, dephosphorylation at Atg13 Ser428/429 is faster than that at Atg1 Thr226 (Fig. 3e, 30–70 min). We thus conclude that Ptc2-mediated dephosphorylation is one mechanism by which phase separation of the PAS is maintained while simultaneously retaining a subpopulation of Atg1 in an activated state.

Dynamic structure of scaffold droplets

We next performed structural analysis of the scaffold droplets using high-speed atomic force microscopy (HS-AFM). Molecules with an S-shape, a distinctive feature of Atg17²³, are distributed irregularly

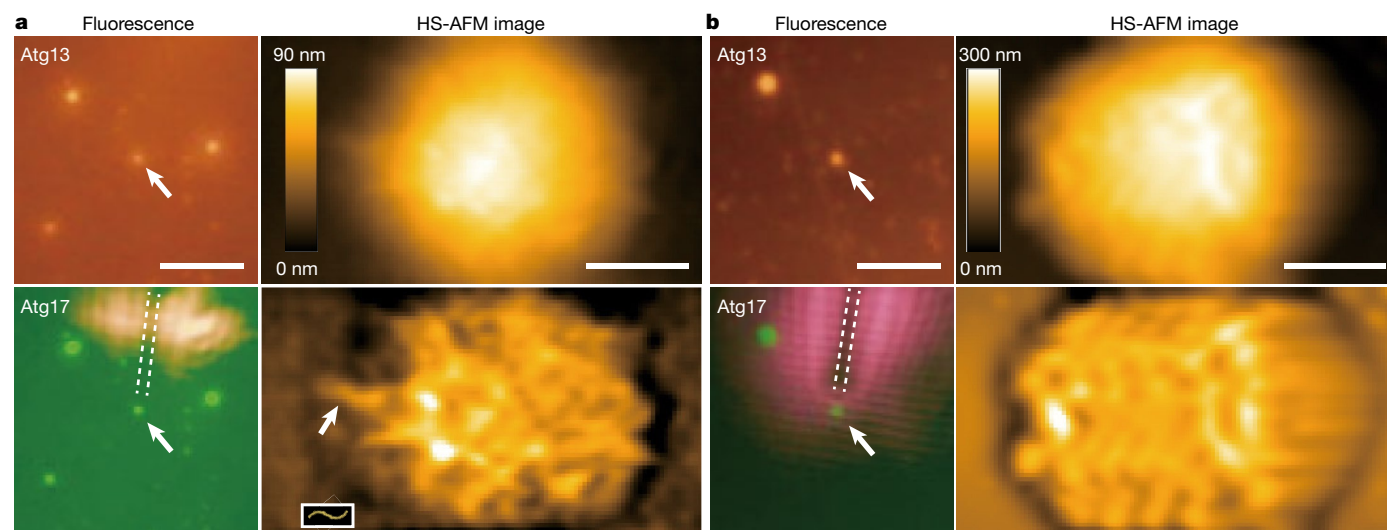


Fig. 4 | The surface structure of scaffold droplets is irregular and dynamic. **a, b,** Visualization of S-shaped Atg17 molecules in scaffold droplets on non-coated (**a**) and 3-aminopropyltriethoxysilane-coated (**b**) coverslips using HS-AFM in combination with fluorescence microscopy. Broken lines indicate the position of the cantilever used for HS-AFM observation. Inset shows Atg17

dimer structure (PDB 5JHF) in the same scale. Bottom HS-AFM images are after fast Fourier transform-bandpass filter. Experiments were repeated independently three times with similar results (**a, b**). **a, b,** Scale bars, 5 μ m (left), 100 nm (**a**, right) and 200 nm (**b**, right).

throughout these droplets, confirming that droplets are not regularly structured (Fig. 4a). S-shaped Atg17 molecules exhibited dynamic behaviour in a restricted area within the droplets (Supplementary Video 7), providing further evidence that the droplets are in a liquid-like state. By contrast, Atg17 was arranged in a regular pattern and showed little movement in droplets that were loaded onto a positively charged coverslip (Fig. 4b, Supplementary Video 8), suggesting that the droplets can mature to a static, solid-like structure, depending on the environment. When FRAP experiments were performed on scaffold droplets, we observed that photobleaching impaired coalescence of droplets and that Atg17 fluorescence rarely recovered (Extended Data Fig. 4). Collectively, these observations indicate that the droplets are liquid-like and randomly structured, and that the droplets eventually mature into a static, ordered state in vitro, as has been observed in other biomolecular condensates that transition from liquid droplets to solid-like states such as gels, glasses and amyloids²⁴.

In vitro reconstitution of the early PAS

PAS formation occurs on the vacuolar membrane. Previous studies indicate that the autophagy-related vacuolar membrane protein Vac8 interacts directly with Atg13²⁵. Monitoring of GFP–Atg1, Atg5–GFP and GFP–Atg8 during starvation showed that puncta containing these proteins were mostly attached to the vacuolar membrane in wild-type cells, whereas about half were detached from the vacuolar membrane in *vac8Δ* cells (Fig. 5a, b, Extended Data Fig. 5a, b). This suggests that Vac8 is at least partly responsible for tethering the PAS to the vacuolar membrane. We next investigated whether Atg1-complex droplets could also be tethered to membranes via Vac8 using giant unilamellar vesicles (GUVs). Atg1-complex droplets were tethered to Vac8-anchored GUVs but not to GUVs lacking Vac8 (Fig. 5c, d, Extended Data Fig. 5c); Atg1 was dispensable for droplet binding to Vac8 GUVs (Fig. 5e). The scaffold droplets were only rarely tethered to GUVs in the presence of the Vac8 mutant that lost the affinity with Atg13²⁶, further confirming that tethering occurs through a specific Atg13–Vac8 interaction (Fig. 5f, Extended Data Fig. 5d, e). The number of droplets tethered to wild-type Vac8 GUVs was rapidly reduced, while their size increased through coalescence events (Fig. 5g, h, Extended Data Fig. 5f, Supplementary Video 9). This result is consistent with our in vivo and in vitro findings (Fig. 1f, g, 2c),

revealing that the droplets remain in a liquid-like state, even in GUVs. FRAP experiments revealed that Atg1-complex droplets tethered to the membrane exchanged 60–100% of their constituent Atg1 molecules within 3 min (Extended Data Fig. 5g), providing further evidence for the liquid-like nature of droplets, even when on membranes. On the basis of these observations, we conclude that a structure similar to the early PAS was reconstituted in vitro using purified proteins and synthetic liposomes, demonstrating that the early PAS is a liquid-like condensate that is tethered to the vacuolar membrane through a specific protein–protein interaction.

Discussion

The relationship between biomolecular condensates and autophagy is generally thought of in passive terms: condensates are targeted for degradation by autophagy⁵. Our results challenge this notion, instead suggesting that the PAS—the central driver of the autophagy mechanism—is a liquid-like biomolecular condensate (summarized in Extended Data Fig. 6). Phase separation is implicated at a fundamental level in PAS formation, with the interactions between IDR-containing Atg1-complex components critical for the early PAS. In previous work, extensive structural analyses have been performed on the Atg1 complex that have established the stoichiometry of Atg1–Atg13, Atg13–Atg17 and Atg17–Atg29–Atg31 interactions^{16,23}. However, in these studies, the majority of IDRs were removed for technical reasons, preventing the observation of Atg1-complex phase-separation events. Here we reveal that cross-linking of Atg17 dimers by the long IDR of Atg13 is the main mechanism of phase separation, a finding supported by a previous structural study describing specific and multivalent Atg13–Atg17 interactions⁶. The resulting liquidity of the PAS is critical for its function in dynamic recruitment of Atg proteins throughout autophagosome formation: for example, liquidity probably results in the concentration and activation of Atg1 kinase for autophagy initiation, and would facilitate the incorporation of Atg9 vesicles, the initial source of autophagosomal membranes⁸, in a manner reminiscent of liquid-phase synapsin clustering of vesicles at synapses²⁷. For these functions, the liquidity of the PAS—which is easily lost by maturation, as observed in vitro (Fig. 4)—is maintained in cells through formation and dissolution events that are mediated by a combination of kinases and phosphatases.

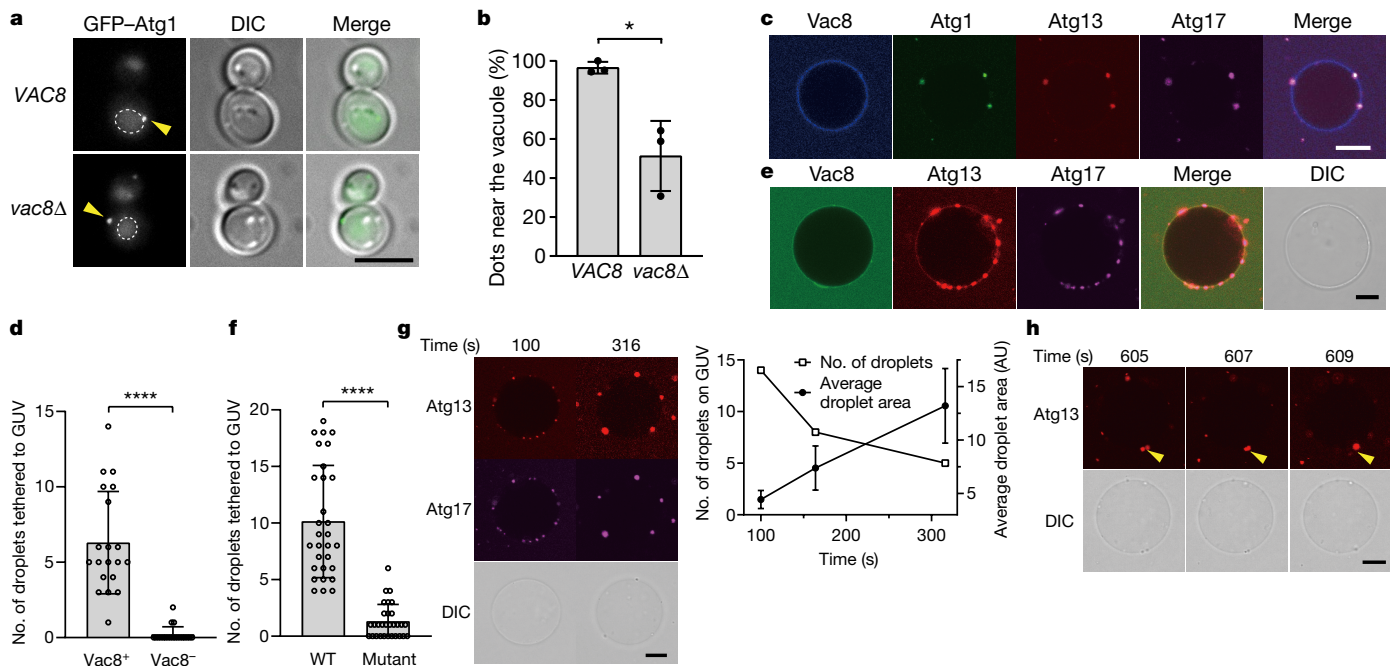


Fig. 5 | The liquid-like PAS is tethered to the vacuolar membrane via specific protein-protein interactions. a, Observation of the PAS by monitoring GFP-Atg1 in wild-type and *vac8Δ* cells treated with rapamycin for 3 h. Broken circles indicate the vacuole. Scale bar, 5 μ m. **b**, Proportion of GFP-Atg1 dots observed adjacent to the vacuole. Data are mean \pm s.d. ($n = 3$ (**b**), 20 (**d**) and 30 (**f**) independent experiments). $^{*}P = 0.0126$, two-sided t -test. **c**, Tethering of Atg1-complex droplets to a Vac8-anchored GUV. **d**, Quantification of the number of Atg1-complex droplets tethered to GUVs after equilibration. $^{****}P = 1.3 \times 10^{-9}$,

two-sided t -test. **e**, Tethering of scaffold droplets to a Vac8-anchored GUV. **f**, Quantification of the number of scaffold droplets tethered to GUVs after equilibration. $^{****}P = 3.6 \times 10^{-13}$, two-sided t -test. **g**, Top, time-dependent change in the number and size of scaffold droplets on a Vac8 GUV. Bottom, graph shows the number and mean area of droplets \pm s.d. ($n =$ droplet numbers). **h**, Coalescence between droplets observed in **g**. Scale bars, 10 μ m unless otherwise indicated. Experiments were repeated independently 3 (**a**), 20 (**c**), 30 (**e**) or 5 (**g**, **h**) times with similar results.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1977-6>.

- Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
- Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
- Suzuki, K. et al. The pre-autophagosomal structure organized by concerted functions of APG genes is essential for autophagosome formation. *EMBO J.* **20**, 5971–5981 (2001).
- Kim, J., Huang, W. P., Stromhaug, P. E. & Klionsky, D. J. Convergence of multiple autophagy and cytoplasm to vacuole targeting components to a perivacuolar membrane compartment prior to de novo vesicle formation. *J. Biol. Chem.* **277**, 763–773 (2002).
- Wang, Z. & Zhang, H. Phase separation, transition, and autophagic degradation of proteins in development and pathogenesis. *Trends Cell Biol.* **29**, 417–427 (2019).
- Yamamoto, H. et al. The intrinsically disordered protein Atg13 mediates supramolecular assembly of autophagy initiation complexes. *Dev. Cell* **38**, 86–99 (2016).
- Suzuki, K., Kubota, Y., Sekito, T. & Ohsumi, Y. Hierarchy of Atg proteins in pre-autophagosomal structure organization. *Genes Cells* **12**, 209–218 (2007).
- Yamamoto, H. et al. Atg9 vesicles are an important membrane source during early steps of autophagosome formation. *J. Cell Biol.* **198**, 219–233 (2012).
- Kawamata, T., Kamada, Y., Kabeya, Y., Sekito, T. & Ohsumi, Y. Organization of the pre-autophagosomal structure responsible for autophagosome formation. *Mol. Biol. Cell* **19**, 2039–2050 (2008).
- Cheong, H., Nair, U., Geng, J. & Klionsky, D. J. The Atg1 kinase complex is involved in the regulation of protein recruitment to initiate sequestering vesicle formation for nonspecific autophagy in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **19**, 668–681 (2008).
- Courchaine, E. M., Lu, A. & Neugebauer, K. M. Droplet organelles? *EMBO J.* **35**, 1603–1612 (2016).
- Wei, M. T. et al. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* **9**, 1118–1125 (2017).
- Mahen, R., Jayasekharan, A. D., Barry, N. P. & Venkitaraman, A. R. Continuous polo-like kinase 1 activity regulates diffusion to maintain centrosome self-organization during mitosis. *Proc. Natl Acad. Sci. USA* **108**, 9310–9315 (2011).

- Kroschwald, S., Maharana, S. & Simon, A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* 1–7, <https://doi.org/10.19185/matters.201702000010> (2017).
- Rabouille, C. & Alberti, S. Cell adaptation upon stress: the emerging role of membrane-less compartments. *Curr. Opin. Cell Biol.* **47**, 34–42 (2017).
- Fujioka, Y. et al. Structural basis of starvation-induced assembly of the autophagy initiation complex. *Nat. Struct. Mol. Biol.* **21**, 513–521 (2014).
- Kamada, Y. et al. Tor-mediated induction of autophagy via an Apg1 protein kinase complex. *J. Cell Biol.* **150**, 1507–1513 (2000).
- Yeh, Y. Y., Wrasman, K. & Herman, P. K. Autophosphorylation within the Atg1 activation loop is required for both kinase activity and the induction of autophagy in *Saccharomyces cerevisiae*. *Genetics* **185**, 871–882 (2010).
- Kamber, R. A., Shoemaker, C. J. & Denic, V. Receptor-bound targets of selective autophagy use a scaffold protein to activate the Atg1 kinase. *Mol. Cell* **59**, 372–381 (2015).
- Torggler, R. et al. Two independent pathways within selective autophagy converge to activate Atg1 kinase at the vacuole. *Mol. Cell* **64**, 221–235 (2016).
- Patel, A. et al. ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).
- Memisoglu, G., Eapen, V. V., Yang, Y., Klionsky, D. J. & Haber, J. E. PP2C phosphatases promote autophagy by dephosphorylation of the Atg1 complex. *Proc. Natl Acad. Sci. USA* **116**, 1613–1620 (2019).
- Ragusa, M. J., Stanley, R. E. & Hurley, J. H. Architecture of the Atg17 complex as a scaffold for autophagosome biogenesis. *Cell* **151**, 1501–1512 (2012).
- Woodruff, J. B., Hyman, A. A. & Boke, E. Organization and function of non-dynamic biomolecular condensates. *Trends Biochem. Sci.* **43**, 81–94 (2018).
- Scott, S. V. et al. Apg13p and Vac8p are part of a complex of phosphoproteins that are required for cytoplasm to vacuole targeting. *J. Biol. Chem.* **275**, 25840–25849 (2000).
- Jeong, H. et al. Mechanistic insight into the nucleus-vacuole junction based on the Vac8p–Nvj1p crystal structure. *Proc. Natl Acad. Sci. USA* **114**, E4539–E4548 (2017).
- Milovanovic, D., Wu, Y., Bian, X. & De Camilli, P. A liquid phase of synapsin and lipid vesicles. *Science* **361**, 604–607 (2018).
- Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **11**, 121–122 (2014).
- Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645 (2004).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains and media

S. cerevisiae strains and plasmids used in this study are listed in Supplementary Tables 1 and 2, respectively. Standard protocols were used for yeast manipulation³⁰. Cells were cultured at 30 °C in nutrient-rich SD + CA medium (0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, 0.5% casamino acids, and 2% glucose) supplemented with appropriate nutrients. Autophagy was induced by transferring cells to nitrogen-starvation SD(–N) medium (0.17% yeast nitrogen base without amino acids and ammonium sulfate, and 2% glucose) or by treating cells with 0.5 µg ml^{–1} rapamycin (Sigma-Aldrich). Treatment of yeast cells with 1,6-hexanediol was performed by adding 1,6-hexanediol and digitonin at a final concentration of 10% and 10 µg ml^{–1}, respectively, to the medium¹⁴. For replenishing nutrients to starved cells, an equal volume of 2× SD + CA medium (doubled concentration of each constituent) was added to culture media.

For galactose induction of GFP–Atg13 protein using the *GAL1* promoter, cells were cultured at 30 °C in nutrient-rich SD + CA medium (0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, 0.5% casamino acids, 2% raffinose and 0.1% glucose) supplemented with appropriate nutrients. The following day, the cell culture (OD₆₀₀ = 1.0–2.0) was supplemented with galactose to a final concentration of 2%, and cultured for an additional 6 h. Autophagy was induced by treating cells with 0.5 µg ml^{–1} rapamycin for 10 min.

Construction of expression plasmids

The pRS316-based low-copy plasmid for expression of GFP–Atg1 and GFP–Atg13 in yeasts under the control of the *ATG1* and *GPD* promoters, respectively, were constructed as described previously¹⁶. The pRS426-based multi-copy plasmid for expression of GFP–Atg13 in yeasts under control of the *GPD* promoter and *GAL1* promoter, respectively, were also constructed similarly. To construct coexpression plasmids encoding SNAP-tagged Atg17, Atg29 and hexahistidine (His₆)-tagged Atg31, genes were amplified by PCR and cloned into the pET28a(+) vector (Novagen) for SNAP-tagged Atg17 and the pACYCDuet-1 vector (Novagen) for Atg29 and His₆-tagged Atg31. To construct expression plasmids encoding N-terminal His₆-tagged and C-terminal SNAP–Twin-Strep-tagged Atg13, genes were amplified by PCR and cloned into the pET11a vector. To construct expression plasmids encoding N-terminal glutathione-S-transferase (GST)-tagged and C-terminal SNAP-tagged Vac8, genes were amplified by PCR and cloned into the pGEX6p-1 vector. To construct expression plasmids encoding N-terminal GST-tagged Ptc2, genes were amplified by PCR and cloned into the pGEX6p-1 vector. To construct expression plasmids encoding the N-terminal SNAP-tagged Atg1 with a HRV3C protease site followed by Flag and His₆ tags, the SNAP-tag gene was amplified by PCR and cloned into the pFastBac Dual-based Atg1 expression vector⁶. The NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs) was used for cloning. Mutations to generate the indicated amino acid substitutions were introduced by PCR-mediated site-directed mutagenesis. All constructs were sequenced to confirm accuracy of cloning.

Protein Expression and Purification

E. coli strain BL21(DE3) cells were used for expression of all recombinant proteins except Atg1. His₆-tagged Atg31 was coexpressed with SNAP–Atg17 and Atg29. After cell lysis, the SNAP–Atg17–Atg29–Atg31 complex was purified by affinity chromatography using a Ni-NTA column (Qiagen). After affinity chromatography, the protein complex was purified on a HiLoad 26/60 Superdex 200 PG column (GE Healthcare) eluted with 20 mM Tris-HCl pH 8.0 and 150 mM NaCl. N-terminal

His₆-tagged and C-terminal SNAP–Twin-Strep-tagged Atg13 was first purified with a Ni-NTA column and then purified using a Strep-TactinXT resin column (IBA Lifesciences). Finally, the proteins were purified on a HiLoad 26/60 Superdex 200 PG column eluted with 20 mM HEPES pH 7.0 and 500 mM NaCl. GST–Vac8–SNAP was first purified using a glutathione–Sephadex 4B (GS4B) column (GE Healthcare). After affinity chromatography, GST was excised using human rhinovirus 3C protease. Vac8–SNAP was again applied to a GS4B column in order to remove the excised GST. Finally, the protein was purified on a HiLoad 26/60 Superdex 200 PG column eluted with 20 mM Tris-HCl, pH 8.0 and 150 mM NaCl. GST–Ptc2 was first purified using a glutathione–Sephadex 4B (GS4B) column (GE Healthcare). After affinity chromatography, GST was excised using human rhinovirus 3C protease. Ptc2 was again applied to a GS4B column in order to remove the excised GST. Recombinant Atg1 was expressed using the baculovirus expression system (Invitrogen) and then purified as described previously⁶. SNAP tag of Atg1, Atg13, and Atg17 was labelled with SNAP–Surface Alexa Fluor 488, SNAP–Surface 549, and SNAP–Surface Alexa Fluor 647 (all from New England Biolabs), respectively, according to the manufacturer's protocol, except for the samples used for AFM experiments, where the SNAP tag of Atg17 was labelled with SNAP–Surface Alexa Fluor 488. For GUV experiments, proteins were dialysed against 20 mM HEPES pH 7.0 and 500 mM NaCl using dialysis tubes, 8 kDa cut-off (GE Healthcare).

FRAP measurements and analysis

For FRAP experiments assessing the PAS, cells treated with rapamycin were imaged on concanavalin A coated glass-bottom dishes (Mattek) to immobilize cells. For FRAP experiments of Atg1-complex droplets attached to a giant liposome, multilamellar liposomes instead of GUVs were used in order to reduce the movement of droplets on the liposome. During FRAP experiments, Atg1 sample was continuously added in the vicinity of the liposome using a micropipette. FRAP experiments were carried out with a FV3000RS confocal laser scanning microscope (Olympus) equipped with an UPLSAPO60XO, NA 1.42 Oil objective (Olympus). For imaging of GFP and SNAP–Surface Alexa Fluor 488 fluorescence, excitation was performed using a 488-nm laser and fluorescence was recorded in a linear sequential mode using a galvano scanner to capture one z-stack for the PAS and 6.9-µm z-stacks with 1.4-µm spacing for the liposome-tethered droplets. Photobleaching was performed using 405-nm and 488-nm laser pulses (1 repeat, 10% intensity, dwell time 5–50 ms) for the PAS and 488-nm laser pulses (1 repeat, 10% intensity, dwell time 45 ms) for the liposome-tethered droplets. Image analysis was carried out with FIJI v.1.52e³¹ or FV31S-SW v.2.1.1.98 (Olympus). For kinetic analysis, relative fluorescence intensity was plotted against time by setting the intensity before quenching as 1.0 and the minimum intensity after quenching as 0.0, and fitted to an exponential recovery curve: $F = A_0(1 - \exp(-k_{\text{off}}t))$ in which A_0 is the maximum recovery at $t = \infty$ and t is time in seconds. This equation was used to determine k_{off} . The $t_{1/2}$ value was calculated as $\ln(2)/k_{\text{off}}$.

Fluorescence correlation spectroscopy

The FCS data shown in Fig. 1 are taken with a TCS SP5 II confocal microscope (Leica). An HC PLAPO 40×/1.10 W CORR CS2 objective lens (Leica) was used for imaging, and the signal was detected with a hybrid detector (HyD, Leica) in a photon-counting mode, with the scanning frequency $f_z = 8,000$ Hz, and the number of pixels $\xi_{\text{max}} = 16$. The size of the detection area was $S_x = 403$ nm, comparable to the size of the diffraction-limited point spread function (PSF) of the observation lens ($w_{xy} = 290$ nm, and $w_z = 1,450$ nm).

Further, we employed scanning FCS (sFCS)³² to minimize photobleaching and the effect of movement of the PAS itself.

In sFCS, a single line scanning data set was taken with a scanning frequency f_z . The intensity of each pixel ξ ($1 \leq \xi \leq \xi_{\text{max}}$) is the integration of fluorescent signals for a pixel dwell time T_x . The illumination volume moves from the leftmost position $\xi = 1$ to the rightmost position ξ_{max} ,

and returns from $\xi = \xi_{\max}$ to 1. The scanner repeats the same procedure every $T_y = 1/f_z$ seconds. For each pixel, the intensity time traces $F_\xi(t_i)$ at time t_i ($i = 0, 1, 2, \dots, 2n - 1$) are used to calculate the temporal autocorrelation curve (where n denotes the number of scans):

$$G_\xi(\tau_\xi) = \frac{\langle \delta F_\xi(t_i) \delta F_\xi(t_i + \tau_\xi) \rangle}{\langle F_\xi(t_i) F_\xi(t_i + \tau_\xi) \rangle}$$

where the angled brackets denote the average over a long time period at each pixel position ξ . The fluorescence of fluctuation is $\delta F_\xi(t_i) = F_\xi(t_i) - \langle F_\xi(t_i) \rangle$ and the lag-time is τ_ξ .

The calculated autocorrelation function $G(\tau)$ was fitted similarly as the conventional fixed point FCS with a 3D one-component model:

$$G(\tau) = \frac{\gamma}{N} \left(1 + \frac{4D\tau}{w_{xy}^2} \right)^{-1} \left(1 + \frac{4D\tau}{w_z^2} \right)^{-\frac{1}{2}}$$

where γ is a geometric factor dependent on the shape of the focal volume (0.35 for 3D Gaussian), N is the number of molecules in the focal volume, as calculated by $N = \langle C \rangle w_{xy}^2 w_z \left(\frac{\pi}{2} \right)^{3/2}$ (where $\langle C \rangle$ denotes the average concentration of molecules in the PSF), D is a diffusion coefficient in $\mu\text{m}^2\text{s}^{-1}$, w_{xy} and w_z are the radial and axial waist of the PSF. We usually calculate the confocal volume as $V_{\text{conf}} = \left(\frac{\pi}{2} \right)^{3/2} w_{xy}^2 w_z$, enabling us to interpret the factors $w_{xy}^2 w_z \pi^{3/2}$ as effective volume V_{eff} , where the particles are actually detected for sources of fluorescence.

However, the original sFCS itself was based on the slow Galvo scanner, which limited its applications to slowly diffusing molecules. We therefore used a Galvo-resonant scanner for high-speed scanning. Furthermore, slowly fluctuating noise was problematic during measurement of living cells due to the movement of subcellular structures. Such noise was statistically removed by the wavelet-based method described in the following section.

Noise removal and data correction for FCS analysis

For correction of photobleaching effects, we averaged the intensity time trace over all pixels $F(t_i) = \langle F_\xi(t_i) \rangle_\xi$ and approximated it as an exponential decay curve $F(t_i) - f(t_i) = f_0 e^{-\frac{t_i}{\tau_b}}$. When we assume that the intensity $F_\xi(t_i)$ obeys a Poisson distribution with mean $\mu_i = f(t_i)$ and variance $\sigma_i^2 = f(t_i)$, we can correct the standard deviation of the intensity and photobleach as $F_\xi^c(t_i) = f(0) + (F_\xi(t_i) - \mu_i)\sigma_0/\sigma_i$ and we have the corrected intensity $F_\xi^c(t_i)$ as in ref.³³:

$$F_\xi^c(t_i) = \frac{F_\xi(t_i)}{\sqrt{f(t_i)/f(0)}} + f(0) \left(1 - \sqrt{\frac{f(t_i)}{f(0)}} \right)$$

For further^{34,35} correction, we applied wavelet-based smoothing to subtract the inhomogeneous intensity fluctuation $F_\xi^W(t_i)$ which is calculated as follows. We applied a wavelet decomposition of the intensity signal $F_\xi^c(t_i)$ which can be achieved by using a scale function ϕ and a wavelet function ψ in multiresolution analysis. In this paper, we used the Haar scaling and wavelet functions as

$$\phi(t) = \begin{cases} 1, & (0 \leq t < 1) \\ 0, & \text{otherwise} \end{cases}$$

$$\psi(t) = \begin{cases} 1, & (0 \leq t < \frac{1}{2}) \\ -1, & (\frac{1}{2} \leq t < 1) \\ 0, & \text{otherwise} \end{cases}$$

If we represent the discrete intensity signal $F_\xi^c(t_i) = \{y_0, y_1, \dots, y_{n-1}\}$ with the length $n = 2^J$ ($J > 0$), the discrete wavelet transform gives the vector of wavelet coefficients with the length n . The coarse approximation of the signal is represented by a linear combination of the shifted scale functions $\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k)$, where j and k are the scale and location of the scale function, respectively. The weights for this function are scale coefficients $c_{j,k}$. The residual details of the signal which are omitted from the coarse approximation can be expressed by a linear combination of the shifted wavelet functions $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$ with the weights $d_{j,k}$. If the coarse approximation level $L < J$ is given, the signal is represented by the wavelet coefficients with 2^L scale and 2^J detail coefficients, $\{c_{L,0}, c_{L,1}, \dots, c_{L,2^L-1}\}$ and $\{d_{j,0}, d_{j,1}, \dots, d_{j,2^j-1}\}$ ($j = L, \dots, J-1$), respectively.

In general, wavelet methods have been known to be advantageous for statistical analyses including the removal of noise when the signal is inhomogeneous in time. For example, when a series of inhomogeneous signal includes a long-term variation and noise, the variation can be captured by large wavelet coefficients and the noise by small coefficients. The successful removal of noise can be achieved by appropriate threshold selection. However, in FCS measurements for living cells, a long variation is usually caused by cell or organelle movement and by noise arising from intensity fluctuation caused by molecular diffusion within a confocal volume. Therefore, we use the term 'fluctuation signal' for such 'noise'. To derive the fluctuation signals of molecular diffusion, we first applied discrete wavelet transform of the original signal $F_\xi^c(t_i) = \{y_0, y_1, \dots, y_{n-1}\}$, and removed noise (fluctuation signals) by setting the detail coefficients $d_{j,k}$ to $\delta_{\lambda_j}^S(d_{j,k})$ or $\delta_{\lambda_j}^H(d_{j,k})$, where $\delta_{\lambda_j}^{S(H)}$ are the soft or hard thresholding functions, respectively,

$$\delta_{\lambda_j}^S(x) = \begin{cases} x - \lambda_j, & (x > \lambda_j) \\ 0, & (|x| \leq \lambda_j) \\ x + \lambda_j, & (x < -\lambda_j) \end{cases}$$

$$\delta_{\lambda_j}^H(x) = \begin{cases} x, & (|x| > \lambda_j) \\ 0, & \text{otherwise} \end{cases}$$

Because the intensity signals obey Poisson distribution, the values of λ_j are level-dependent thresholds for Poisson noise, and the translation-invariant Poisson smoothing using Haar wavelets (TIPSH) algorithm was used³⁵.

Second, we applied inverse wavelet transform of the wavelet coefficients and to determine the estimated long-term variation $F_\xi^W(t_i) = \{\hat{y}_0, \hat{y}_1, \dots, \hat{y}_{n-1}\}$. The fluctuation signals are calculated by subtracting this variation from the original inhomogeneous signals ($F_\xi^c(t_i) - F_\xi^W(t_i)$). Finally, the absolute values of fluctuation signals can be calculated by adding the time average of the variation $\overline{F_\xi^W(t_i)}$, giving

$$F_\xi^{CW}(t_i) = \overline{F_\xi^W(t_i)} + (F_\xi^c(t_i) - F_\xi^W(t_i))$$

We have implemented the above method with the Python programming language (Python Software Foundation, Python v.3.6)³⁶. The autocorrelation function after noise removal was calculated by a multiple-tau correlation algorithm³⁷, followed by a nonlinear least-squares fitting to the theoretical model. The radial and axial waist of the PSF were determined by measuring fluorescent solutions and fixed values on fitting.

Reconstitution and microscopy of protein-rich droplets

Liquid droplets of the Atg13-Atg17-Atg29-Atg31 complex and the Atg1 complex were formed by dilution of proteins from a stock solution into buffer as follows; In Fig. 2b, liquid droplets of 3.3 μM Atg1 complex (SNAP-Atg1(D211A), Atg13-SNAP, SNAP-Atg17-Atg29-Atg31) were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM HEPES, pH 7.0, 250 mM NaCl) with subsequent incubation at 25 °C for the indicated times. In Fig. 2c, liquid droplets of

Article

7.5 μM Atg1 complex (Atg1(D211A), Atg13–SNAP, SNAP–Atg17–Atg29–Atg31) were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM Bis-Tris-HCl, pH 5.5, 400 mM NaCl) at 25 °C. In Fig. 2e, 4.8 μM Atg1 complexes (SNAP–Atg1(D211A), Atg13–SNAP, SNAP–Atg17–Atg29–Atg31) were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM MES, pH 6.0, 300 mM NaCl) with subsequent incubation for 30 min at 25 °C. In Fig. 2g and Extended Data Fig. 2h, liquid droplets of 6 μM Atg13–Atg17–Atg29–Atg31 containing the indicated mutations were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM HEPES, pH 7.0, 250 mM NaCl) with subsequent incubation at 25 °C for the indicated time. In Extended Data Fig. 2d, liquid droplets of 6 μM Atg13–Atg17–Atg29–Atg31 were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM HEPES, pH 7.0, 250 mM NaCl) with subsequent incubation for 10 min at 25 °C. Next, 1,6-hexanediol or buffer was added at a final concentration of 5% to liquid droplets at 25 °C. In Extended Data Fig. 2i, liquid droplets of the indicated concentration of Atg13–Atg17–Atg29–Atg31 were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM MES, pH 6.0, 400 mM NaCl) with subsequent incubation for 3 min at 25 °C. Phase separation was scored ‘droplet’ or ‘no droplet’, depending on the presence or absence of protein droplets. In Fig. 3b, liquid droplets of 2.7 μM Atg13–Atg17–Atg29–Atg31 containing either phosphorylated Atg13 or non-phosphorylated Atg13 were formed by dilution of proteins from a stock solution into buffer (final concentrations: 50 mM HEPES, pH 7.0, 250 mM NaCl) with subsequent incubation at 25 °C for the indicated times.

These samples were mixed in a microtube and imaged on a glass-bottom dish (Mattek) coated with 3% bovine serum albumin (BSA) (Wako). A FV3000RS was used for fluorescence imaging. 488-nm, 561-nm, and 640-nm lasers were used for excitation of Atg1 labelled with SNAP-Surface Alexa Fluor 488, Atg13 labelled with SNAP-Surface 549, and Atg17 labelled with SNAP-Surface Alexa Fluor 647, respectively, and fluorescence was recorded in linear sequential mode using a galvano scanner. Quantitative analysis was carried out with Fiji.

In vitro pull-down assay

Purified proteins were incubated with GST-accept beads (Nacalai Tesque) at 4 °C for 30 min. After the beads were washed three times with PBS, proteins were eluted by 10 mM glutathione in 50 mM Tris-HCl (pH 8.0). The samples were separated by SDS–PAGE. Protein bands were detected by One Step CBB (BIO CRAFT).

Phosphorylation of Atg13 by TORC1

Purification of TORC1 from yeast was performed as previously reported¹⁷. In the final step of purification, TORC1 was eluted with elution buffer (250 ng μl^{-1} Flag peptide (Sigma, F3290), 100 mM NaCl, 31 mM Tris-HCl, pH 7.5). Atg13 (3 μM) was phosphorylated by TORC1 in reaction buffer (1 mM ATP, 1 mM MgCl_2 , 1 \times protease inhibitor cocktail (Nacalai, 03969-21), 1 mM PMSF, 100 mM NaCl, 50 mM Tris-HCl, pH 7.5) for 17 h at 20 °C. After the reaction, NaCl was added to a final concentration of 500 mM. Dilution with buffer (20 mM HEPES, pH 7.0, 500 mM NaCl) and concentration were repeated for phosphorylated Atg13 to exchange the buffer. Samples were separated by SDS–PAGE and Zn^{2+} -Phos-tag SDS–PAGE (Wako). Zn^{2+} -Phos-tag SDS–PAGE was performed using 20 μM Phos-tag solution. Protein bands were detected by One Step CBB. For western blotting, protein bands were detected by C-DiGit Blot Scanner (LI-COR Biotechnology).

Phosphorylation of the Atg1 complex

For Extended Data Fig. 3f, liquid droplets of 3.3 μM Atg1 complex containing either wild-type Atg1 or Atg1(D211A) were formed by dilution of the protein from a stock solution into buffer (final concentration: 50 mM HEPES, pH 7.0, 250 mM NaCl) and then incubated for 10 min at 25 °C. Next, one-tenth volume of ATP solution (10 mM ATP, 10 mM

MgCl_2 , 10 mM HEPES, pH 7.0, 250 mM NaCl) was added to liquid droplets at 25 °C. Samples were then collected at the indicated time points. The samples were separated by SDS–PAGE and detected by One Step CBB. Fluorescence of the same samples was imaged using FV3000RS on 3% BSA coated glass-bottom dishes. Turbidity was measured by sample optical density at 350 nm using a NanoDrop 2000 (Thermo Scientific).

Autophosphorylation assays

For confirming the specificity of anti-T226-P antibodies (Extended Data Fig. 3b), ATP (–) samples were prepared by incubating the Atg1 complex containing either wild-type or T226A of SNAP–Atg1–Flag–His₆ in 50 mM MES, pH 6.0, 150 mM NaCl for 30 min at 25 °C. ATP (+) samples were prepared by incubating the same protein complex in 50 mM MES, pH 6.0, 150 mM NaCl for 30 min at 25 °C followed by incubation with 1 mM ATP-Mg in the same buffer solution for 30 min at 25 °C. The samples were separated by SDS–PAGE and subjected to western blotting using anti-T226-P and anti-Flag antibodies. For Fig. 3c, liquid droplets of 0.2 μM Atg1 complex containing either wild-type Atg13 or Atg13(F430A) were formed by incubation in 50 mM MES, pH 6.0, 300 mM NaCl for 30 min at 25 °C. A sample comprising Atg1 alone was also prepared in the same buffer. Next, the ATP-Mg solution was added to samples at a final concentration of 1 mM at 25 °C. Samples were then collected at indicated time points. Samples were separated by SDS–PAGE and subjected to western blotting. Protein bands were detected by C-DiGit Blot Scanner (LI-COR Biotechnology). Quantitative analyses were carried out using Fiji.

Dephosphorylation of the Atg1 complex

For Fig. 3d, e, liquid droplets of 2 μM Atg1 complex were formed by dilution of the protein from a stock solution into buffer (final concentration: 50 mM HEPES, pH 7.0, 250 mM NaCl) with subsequent incubation for 10 min at 30 °C. Next, the solution containing Atg1-complex droplets was supplemented with one-tenth volume of the ATP solution (10 mM ATP, 10 mM MgCl_2 , 10 mM HEPES, pH 7.0 and 250 mM NaCl) and incubated for 20 min at 30 °C. Finally, the solution was supplemented with Ptc2 and MnCl_2 to final concentrations of 1 μM and 3 mM, respectively, and incubated for 40 min at 30 °C. Samples were collected at the indicated time points for further analysis.

Antibodies

Polyclonal antibodies against *S. cerevisiae* Atg1 with phosphorylation at Thr226 (anti-T226-P antibody) were raised according to a previous report¹⁸. Antibodies were raised in rabbits against a phosphorylated peptide, FLPNTSLAE[pThr]LCGSPLY, which corresponds to the sequence of the activation loop of Atg1. ELISA was performed using the peptides with and without phosphorylation at Thr226 to confirm the specificity of antibodies against the phosphorylated peptide. The peptide synthesis, antibody generation, ELISA, and purification were performed by GenScript. Polyclonal antibodies against *S. cerevisiae* Atg13 with phosphorylation at Ser428 and Ser429 (anti-S428/9-P antibody) were generated as described in a previous study¹⁶. Anti-Flag M2 antibody was purchased from Sigma (F3165). Anti-HA antibody was purchased from MBL (M180-3S). Anti-mouse IgG (Fab specific)–peroxidase antibody produced in goat was purchased from Sigma (A9917). Anti-rabbit IgG (whole molecule)–peroxidase antibody produced in goat was purchased from SIGMA (A6154).

Sample preparation for HS-AFM observation

For HS-AFM imaging, coverslips (24 \times 32 mm, 0.13–0.17 mm thick) (Matsunami Glass) were used as a solid support. Coverslips were immersed in 5 M KOH solution for 1 h, followed by three washes in Milli-Q water. Cleaned coverslips were subsequently sonicated in Milli-Q water for 20 min and stored in ethanol at 4 °C until use. Ethanol was completely eliminated before each experiment. Atg13–SNAP (1 μM) and SNAP–Atg17–Atg29–Atg31 (1 μM) were mixed in 20 μl of observation

buffer (250 mM NaCl, 20 mM HEPES-NaOH, pH 7.0) and deposited onto cleaned coverslips. After a 5-min incubation, excess proteins were washed out with observation buffer. For observation of data presented in Fig. 4b and Supplementary Video 8, the coverslip was treated with 0.02% 3-aminopropyltriethoxysilane for 5 min and then washed in MilliQ-water before deposition of the protein mixture.

HS-AFM imaging

HS-AFM images were acquired in tapping mode using a tip-scan type HS-AFM instrument³⁸ (Nano Explorer PS-NEX, Research Institute of Biomolecule Metrology Co.) equipped with a fluorescence microscope. We used cantilevers measuring $\sim 9\ \mu\text{m}$ long, $\sim 2\ \mu\text{m}$ wide and $\sim 0.13\ \mu\text{m}$ thick with a resonant frequency of $\sim 1.5\ \text{MHz}$ and a spring constant of $0.1\text{--}0.2\ \text{N m}^{-1}$ (BL-AC10DS, Olympus). Scaffold droplets were selected for observation by fluorescence imaging using SNAP-Surface 549 labelled to Atg13–SNAP and Alexa Fluor 488 labelled to SNAP–Atg17 and located to the HS-AFM scanning area ($\sim 4 \times \sim 6\ \mu\text{m}^2$) before nanoscale imaging with HS-AFM. HS-AFM imaging conditions were as follows: scan size, $500 \times 250\ \text{nm}^2$ (Fig. 4a and Supplementary Video 7) or $1000 \times 500\ \text{nm}^2$ (Fig. 4b and Supplementary Video 8); pixel size, 120×60 pixels; imaging rate, ~ 3.1 frames per s (fps). All imaging was performed at $23\ ^\circ\text{C}$. IGOR Pro (WaveMetrics) based software for HS-AFM was used to process the image, using features such as Gaussian filtering, automatic flattening and fast Fourier transform-frequency filtering³⁹.

Preparation of GUVs

The natural swelling method is used to prepare GUVs containing PEMCC in a buffer from a dry lipid film⁴⁰. To prepare GUVs using electrically neutral lipids in a buffer of normal physiological ion concentration ($\sim 150\ \text{mM NaCl}$), we used a very small fraction of PEG-lipid (that is, the PEG-lipid method)^{40,41}. We prepared $200\ \mu\text{l}$ of a $1\ \text{mM}$ phospholipid mixture consisting of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE), 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[4-(*p*-maleimidomethyl)cyclohexane-carboxamide] (PEMCC), and 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[methoxy(polyethylene glycol)-2000 (DPPE-PEG2000)] (all lipids were purchased from Avanti Polar Lipids) at a molar ratio of 69:20:10:1 in chloroform in a 5-ml glass vial, and then produced a homogeneous thin film of lipid mixture by evaporation of chloroform using the gentle application of nitrogen gas. For the complete removal of chloroform, we then placed the glass vial in a vacuum desiccator connected to a rotary pump overnight. The following day, we prehydrated the thin lipid film on the bottom of the glass vial using $20\ \mu\text{l}$ of water at $60\ ^\circ\text{C}$ for 7 min. Thereafter $1\ \text{ml}$ of HEPES buffer (20 mM HEPES, pH 7.0, 150 mM NaCl and 1 mM EGTA) containing $0.1\ \text{M}$ sucrose was added and samples were incubated for 2–3 h at $60\ ^\circ\text{C}$.

Observation of GUVs

For dilution, $200\ \mu\text{l}$ of GUV solution was added to $800\ \mu\text{l}$ of HEPES containing $0.1\ \text{M}$ glucose solution (external solution) into a hand-made microchamber that was formed on a glass slide (Matsunami Glass) by depositing (in parallel) two bar-shaped silicon-rubber (3-mm silicon sheet) spacers between a cover slip (Micro cover glass, Muto Pure Chemicals) and the glass slide⁴². The microchamber was coated with 0.10% (w/v) BSA prepared in the same buffer used in experiments to avoid strong contact of GUVs with the glass surface. To increase the contrast of the GUVs for DIC observation, the interior and exterior of GUVs were filled with $0.1\ \text{M}$ sucrose or $0.1\ \text{M}$ glucose, respectively. Observation of GUVs was performed using a FV3000RS as described above at room temperature ($\sim 23\ ^\circ\text{C}$).

Single GUV method for observation of GUV–protein interactions

The single-giant unilamellar vesicle (GUV) method, by which proteins can be precisely added one by one in the vicinity of a GUV, was used for

observation of GUV–protein interactions⁴³. Purified proteins in HEPES buffer containing $0.1\ \text{M}$ glucose were added slowly one by one into the vicinity of a single GUV through a $12\text{--}15\ \mu\text{m}$ diameter glass micropipette, the position of which was controlled by a micromanipulator (Narishige) at room temperature⁴². The distance between the GUV and the tip of the micropipette was maintained at $\sim 50\ \mu\text{m}$. The glass micropipette was prepared as follows: first we pulled a glass tube of 1.0-mm diameter to a needle point using a puller, and the needle point was then microforged to the desired tip diameter (all equipment from Narishige). Proteins in the external solution of GUVs were filled in the micropipette by aspiration using a vacuum pump (ULVAC KIKO), and then the micropipette was held by a micromanipulator, enabling us fine control over tip positioning in the vicinity of the GUV. Protein application pressure in the vicinity of the GUV was controlled by changing the height of a vertical column of water to which the micropipette was hydraulically connected⁴³. The application pressure was measured using a differential pressure transducer (Validyne), pressure amplifier (Karone), and a digital multimeter. For the constant application of proteins in the vicinity of GUVs using the micropipette, we first determined the equilibrium pressure by bringing the tip of the micropipette near to a small vesicle and adjusting the pressure to keep the vesicle at the tip of the micropipette. After fixing equilibrium pressures, an additional $\sim 300\ \text{mV}$ pressure was added to constantly apply protein solution in the vicinity of GUVs.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All relevant data are available from the authors. Source data for gels and blots are provided as Supplementary Information. Source Data for graphs are provided with the paper.

- Adams, A., Gottschling, D. E., Kaiser, C. A. & Stearns, T. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, 1998).
- Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
- Ries, J. & Schwille, P. Studying slow membrane dynamics with continuous wave scanning fluorescence correlation spectroscopy. *Biophys. J.* **91**, 1915–1924 (2006).
- Ries, J., Chiantia, S. & Schwille, P. Accurate determination of membrane dynamics with line-scan FCS. *Biophys. J.* **96**, 1999–2008 (2009).
- Kolaczky, E. D. & Dixon, D. D. Non-parametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics. *Astrophys. J.* **534**, 490–505 (2000).
- Kolaczky, E. D. Non-parametric estimation of gamma-ray burst intensities using Haar wavelets. *Astrophys. J.* **483**, 340–349 (1997).
- Rossum, G. V. *Python tutorial. Technical Report CS-R9526* (Centrum voor Wiskunde en Informatica, 1995).
- Magatti, D. & Ferri, F. Fast multi-tau real-time software correlator for dynamic light scattering. *Appl. Opt.* **40**, 4011–4021 (2001).
- Fukuda, S. et al. High-speed atomic force microscope combined with single-molecule fluorescence microscope. *Rev. Sci. Instrum.* **84**, 073706 (2013).
- Imamura, M. et al. Probing structural dynamics of an artificial protein cage using high-speed atomic force microscopy. *Nano Lett.* **15**, 1331–1335 (2015).
- Alam, J. M., Kobayashi, T. & Yamazaki, M. The single-giant unilamellar vesicle method reveals lysenin-induced pore formation in lipid membranes containing sphingomyelin. *Biochemistry* **51**, 5160–5172 (2012).
- Alam, J. M. & Yamazaki, M. Spontaneous insertion of lipopolysaccharide into lipid membranes from aqueous solution. *Chem. Phys. Lipids* **164**, 166–174 (2011).
- Karal, M. A., Alam, J. M., Takahashi, T., Levadny, V. & Yamazaki, M. Stretch-activated pore of the antimicrobial peptide, magainin 2. *Langmuir* **31**, 3391–3401 (2015).
- Yamazaki, M. The single GUV method to reveal elementary processes of leakage of internal contents from liposomes induced by antimicrobial substances. *Adv. Planar Lipid Bilayers Liposomes* **7**, 121–142 (2008).
- Noda, T., Matsuura, A., Wada, Y. & Ohsumi, Y. Novel system for monitoring autophagy in the yeast *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **210**, 126–132 (1995).

Acknowledgements We thank H. Yamamoto, T. Kawamata, Y. Kamada and D. S. Goldfarb for providing plasmids and strains for yeast experiments, Y. Ishii for assistance with protein preparation and H. Tochio for critical advice. This work was supported in part by JSPS KAKENHI grant number 25111004, 18H03989, 19H05707 (to N.N.N.), 15H01651, 17H05894, 17K07319 (to Y.F.), 19K16344 (to D.N.), 16H06375 (to Y. Ohsumi), 26119003, 17H06121 (to

Article

T.A.), 16H06280, 18H04853 (to K.S.), 19H05795, 19H03394, 16H06280 (to Y. Okada), 18H04751, 19H22520 (to K.M.), JST CREST grant number JPMJCR13M7 (to N.N.N.), JPMJCR15G2 (to Y. Okada) and grants from RIKEN (pioneering project 'Dynamic Structural Biology' to Y. Okada), from the Takeda Science Foundation (to N.N.N. and Y.F.), from Mochida Memorial Foundation for Medical and Pharmaceutical Research (to N.N.N.), from Tokyo Biochemical Research Foundation (to N.N.N. and J.M.A.), and from the Naito Foundation (to N.N.N. and Y.F.).

Author contributions Y.F., R.L.K. and N.N.N. conceived the project and performed initial droplet-formation experiments. Y.F. purified recombinant proteins and performed most of the in vitro experiments. J.M.A. performed GUV experiments. D.N. and T.A. performed HS-AFM observations. Y.F., K.S., Y. Ohsumi and N.N.N. performed yeast experiments. K.M. and Y. Okada

performed FCS analysis. All authors analysed the data. Y.F., A.I.M. and N.N.N. wrote the manuscript with input from all other authors. N.N.N. supervised the work.

Competing interests The authors declare no competing interests.

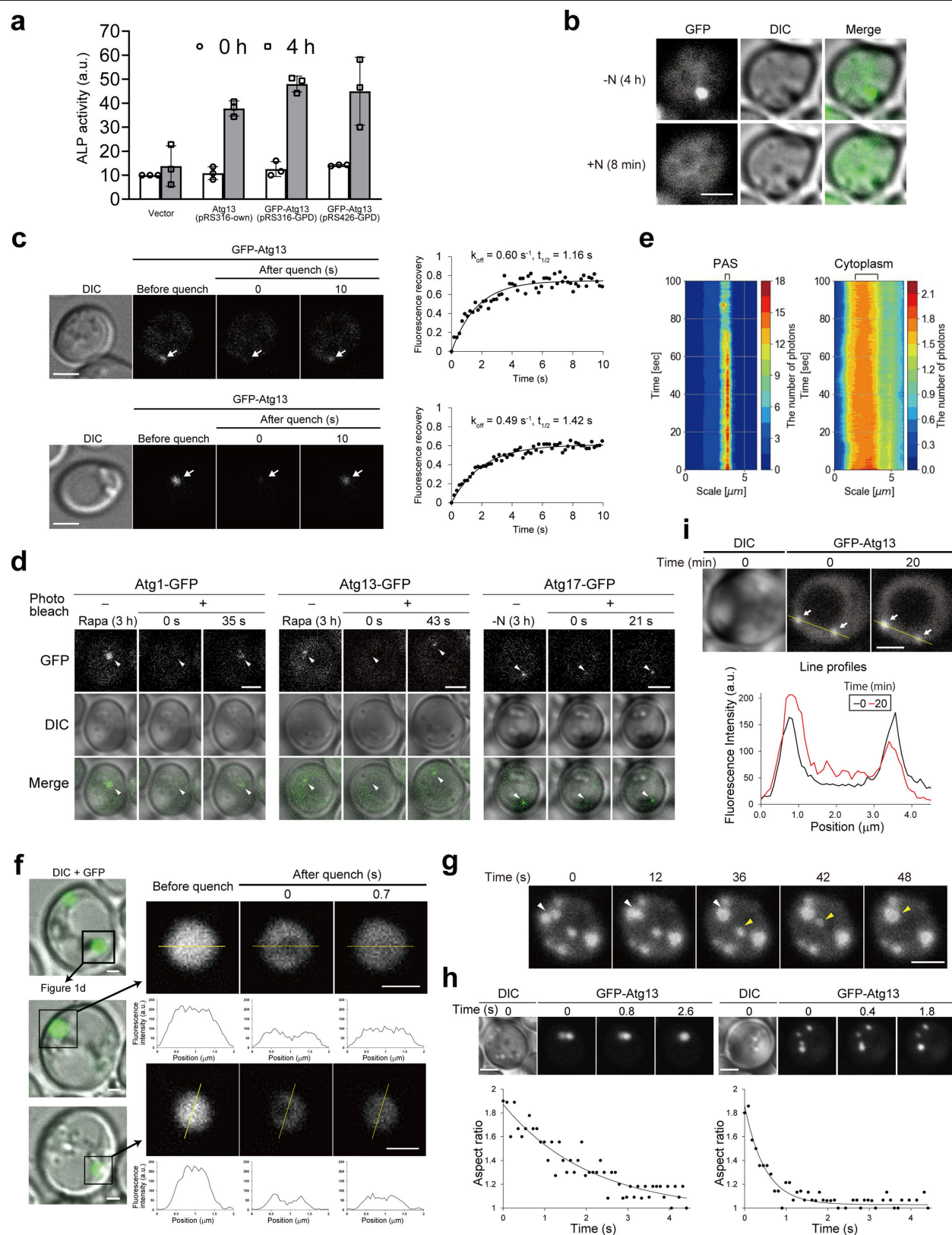
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1977-6>.

Correspondence and requests for materials should be addressed to N.N.N.

Peer review information *Nature* thanks Peter Hinterdorfer, Fulvio Reggiori, Jeffrey Woodruff and Li Yu for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

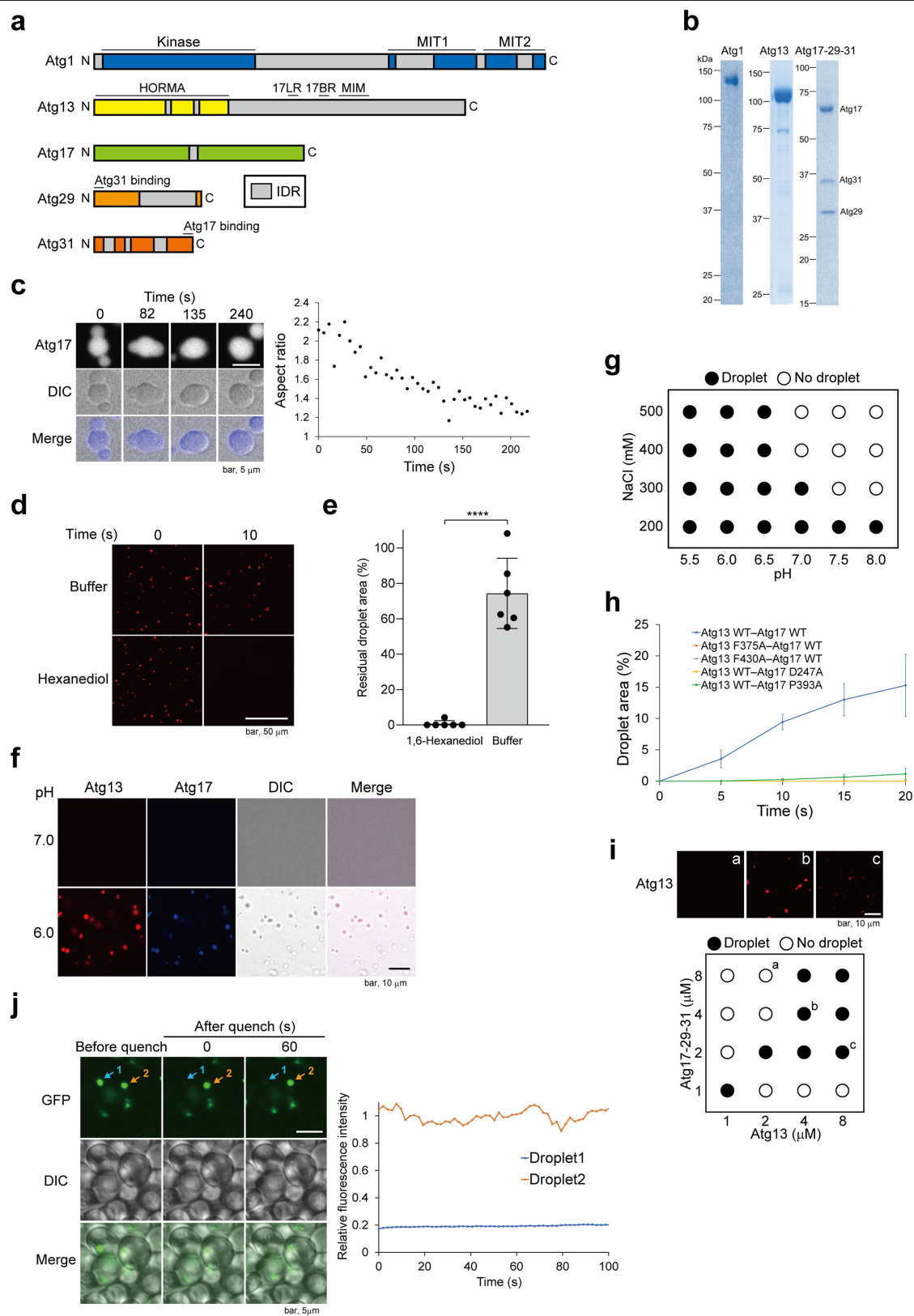


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Characterization of the PAS using fluorescence

microscopy. a, Measurement of alkaline phosphatase activity of yeast cells overexpressing GFP-Atg13 performed based on previous reports⁴⁴. Data are mean \pm s.d. ($n=3$ independent experiments). Bar colour indicates hours after onset of nitrogen starvation. **b**, Dissolution of Atg13-GFP puncta after addition of nitrogen source. Experiments were repeated independently twice with similar results. **c**, Two additional examples of GFP-Atg13 fluorescence recovery, related to Fig. 1a. Experiments were repeated independently three times with similar results, which are shown here and in Fig. 1a. **d**, Fluorescence of endogenously expressed Atg1-GFP, Atg13-GFP and Atg17-GFP puncta rapidly recovers after photobleaching. Rapa and -N indicate rapamycin treatment and nitrogen starvation, respectively. Experiments were repeated independently twice with similar results. **e**, Kymograph of FCS data shown in Fig. 1b. **f**, Two additional examples of partial fluorescence recovery of giant

GFP-Atg13 droplets, related to Fig. 1d. DIC of Fig. 1d experiment is also shown. Experiments were repeated independently three times with similar results, which are shown here and in Fig. 1d. **g**, Coalescence of GFP-Atg13 puncta observed after removing 1,6-hexanediol. The images are the sum of five z-slices of GFP fluorescent images. Experiments were repeated independently three times with similar results. **h**, Two additional examples of the coalescence of two PAS precursors, related to Fig. 1g. Experiments were performed three times with similar results, which are shown here and in Fig. 1g. **i**, An additional example of Ostwald ripening of the PAS, related to Fig. 1h. **i**, The images are the sum of four z-slices of GFP fluorescent images. The bottom graph shows the line profile of fluorescence intensity in the top image. Experiments were repeated independently twice with similar results, which are shown here and in Fig. 1i. Scale bars, 2 μ m, except in **f**, 1 μ m.

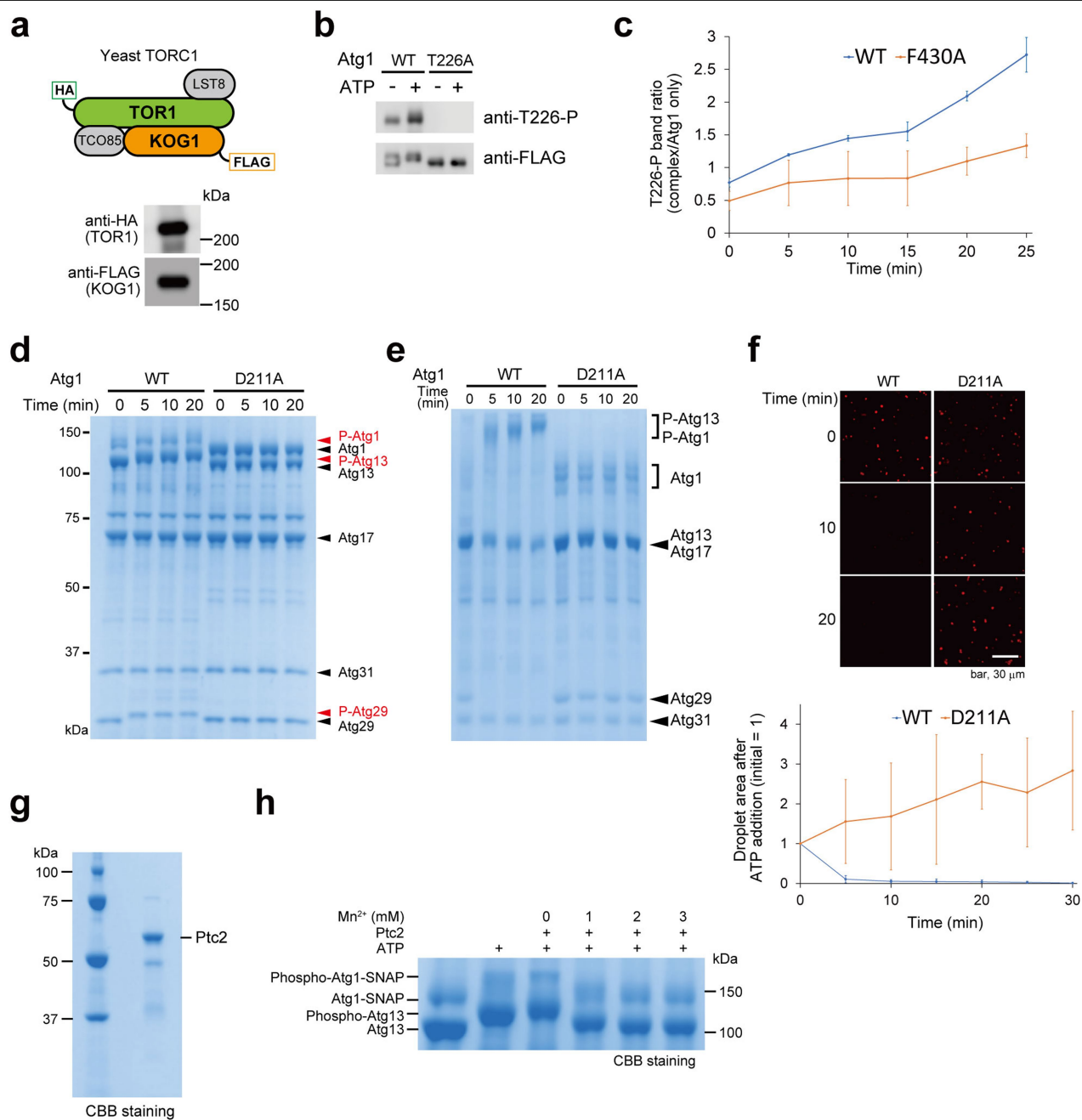


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Characterization of the Atg1-complex droplets

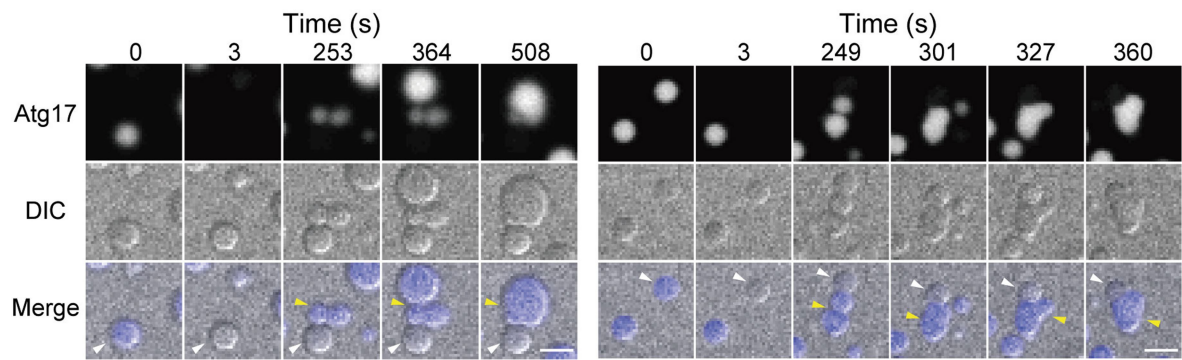
in vitro. **a**, Domain organization of Atg1-complex components. Grey regions indicate IDRs consisting of ten or more residues predicted to be disordered by DISOPRED²⁹. Bar length is approximately proportional to the number of residues. **b**, SDS-PAGE of purified SNAP-tagged proteins used for in vitro analyses. Experiments were repeated independently twice with similar results. For gel source data, see Supplementary Fig. 1. **c**, An additional example of coalescence of Atg1-complex droplets observed in vitro, related to Fig. 2c. The right panel shows the change of the aspect ratio during coalescence. Experiments were repeated independently twice with similar results, which are shown here and in Fig. 2c. **d**, Formation of liquid droplets of the scaffold complex and their dissociation by 1,6-hexanediol treatment. Experiments were repeated independently six times with similar results. **e**, Quantification of the

residual droplet area in **d**. Data are mean \pm s.d. ($n = 6$ independent experiments). **** $P = 3.9 \times 10^{-6}$, two-sided t -test. **f**, The effect of pH on the formation of scaffold droplets. The concentrations of NaCl and Atg13-Atg17-Atg29-Atg31 are 500 mM and 4 μ M, respectively. The experiment was repeated independently three times with similar results. **g**, Phase diagram of the formation of scaffold droplets at indicated NaCl concentrations and pH values. The protein concentration is 4 μ M. Experiment was performed once. **h**, Time-course analysis of droplet area in Fig. 2g. Data are mean \pm s.d. ($n = 3$ independent experiments). **i**, Phase diagram of droplet formation upon mixing of Atg13 and Atg17-Atg29-Atg31 at indicated protein concentrations. Representative images at a, b and c in the diagram are shown above the diagram. Experiment was performed once.



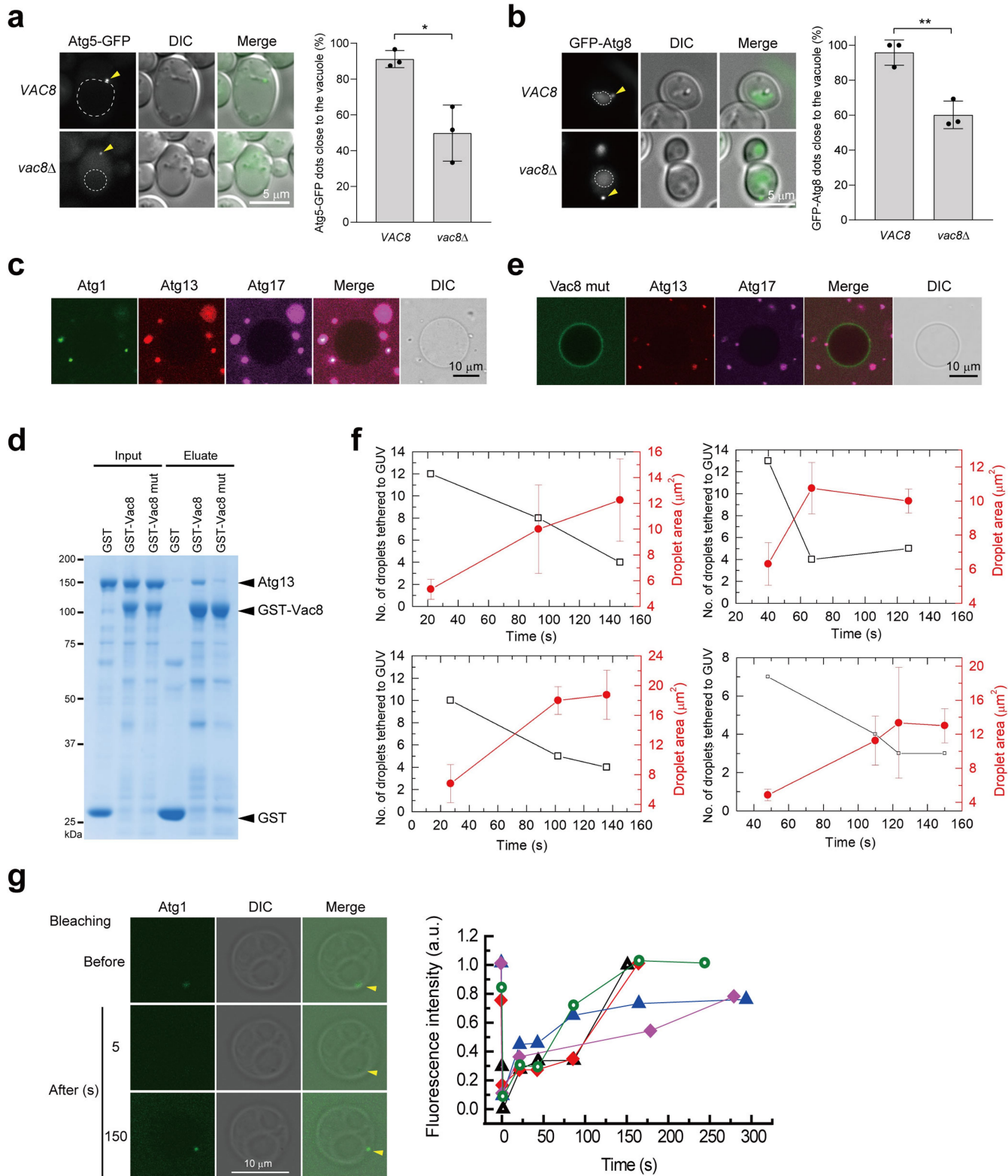
Extended Data Fig. 3 | In vitro assay determination of Atg1 and Ptc2 phosphoregulatory activity. **a**, Purification of TORC1 from yeast. The experiment was repeated independently twice with similar results. **b**, Confirmation of the specificity of the anti-T226-P antibodies. Experiment was performed once. **c**, Quantification of the results in Fig. 3c. Data are mean \pm s.d. ($n=3$ independent experiments). **d**, **e**, Phosphorylation-mediated band shifts of Atg1, Atg13 and Atg29 upon incubation of the Atg1 complex with ATP analysed by conventional (**d**) and Phos-tag SDS-PAGE (**e**). Experiment was

repeated independently three times with similar results. **f**, Effect of Atg1-mediated phosphorylation on Atg1-complex droplets. Bottom graph shows time-course analysis of droplet area. Data are mean \pm s.d. ($n=3$ independent experiments). **g**, SDS-PAGE of the recombinant Ptc2 used in this study. Experiments were repeated independently twice with similar results. **h**, Mn²⁺-dependent dephosphorylation of Atg1 and Atg13 by Ptc2. Experiment was performed once. For gel source data, see Supplementary Fig. 1 (**a**, **b**, **d**, **e**, **g**, **h**).



Extended Data Fig. 4 | In vitro observation of droplet maturation.
Photobleaching experiments were performed for the scaffold droplets in vitro. White and yellow arrow heads indicate photobleached and non-bleached

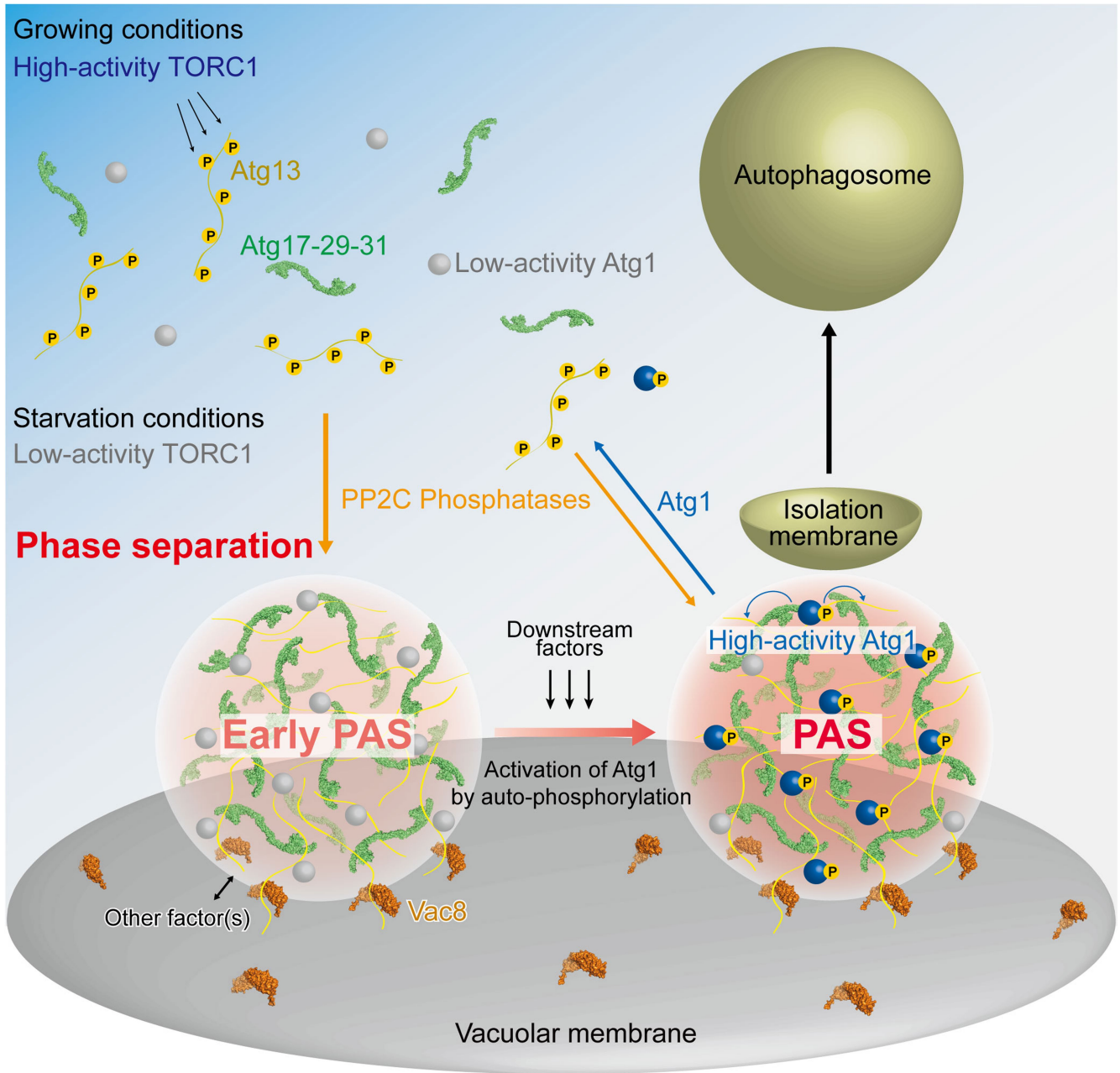
droplets, respectively. Photobleaching was performed between 0 and 3 s. Non-bleached droplets coalesce, whereas bleached droplets do not. Experiments were repeated independently twice with similar results. Scale bars, 5 μ m.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Droplets tethered to GUVs via specific Atg13–Vac8 interaction retain their liquid-like nature. a, b, Deletion of *VAC8* results in mislocalization of Atg5–GFP puncta away from the vacuole (**a**) and GFP–Atg8 puncta (**b**). Atg5–GFP and GFP–Atg8 were observed following 3 or 5 h rapamycin incubation, respectively. Data are mean \pm s.d. ($n=3$ independent experiments). * $P=0.0119$, ** $P=0.0044$, two-sided t -test. **c**, Lack of tethering of Atg1-complex droplets to Vac8-free GUVs. Experiments were repeated independently 20 times with similar results. **d**, Impaired interaction of the Vac8 mutant with Atg13 demonstrated by in vitro pull-down assay. Experiments were repeated independently three times with similar results. For gel source data, see Supplementary Fig. 1. **e**, Near complete lack of tethering of scaffold

droplets to Vac8 mutant-anchored GUVs. Experiments were repeated independently 30 times with similar results. **f**, Additional examples of time-dependent change in the number and size of scaffold droplets on Vac8-GUVs, related to Fig. 5g. The number and average area of droplets \pm s.d. (n = droplet numbers) are shown. Experiments were repeated independently five times with similar results, which are shown here and in Fig. 5g. **g**, FRAP experiments of Atg1–SNAP in Atg1-complex droplets attached to Vac8-anchored multilamellar vesicles. Multilamellar vesicles were used instead of GUVs to reduce the movement of droplets on vesicles. The bottom graph indicates the ratio of fluorescence intensity at each time point in comparison to the initial intensity. Data from seven independent experiments are shown.



Extended Data Fig. 6 | Proposed model of the PAS and its regulation by phosphorylation. Under growing conditions, hyperphosphorylation of Atg13 by TORC1 inhibits Atg1-complex formation and phase separation. Upon starvation, TORC1 activity is inhibited and Atg13 is dephosphorylated by PP2C phosphatases, which leads to Atg1-complex formation. The Atg1 complex then undergoes phase separation to form a liquid droplet (early PAS), which is

tethered to the vacuolar membrane through Atg13–Vac8 interaction. The early PAS activates Atg1 kinase by accelerating autophosphorylation and at the same time recruits downstream Atg factors, thereby transforming into the mature PAS from which isolation membrane is generated. Continuous phosphorylation and dephosphorylation of Atg13 at the PAS would contribute to maintaining the liquid property of the PAS.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Fluorescence and DIC images were obtained by FV31S-SW ver. 2.3.1.163.
AFM images were obtained by FalconTS_1.0.0.
Blot images were obtained by Image Studio Ver. 4.0.21.
SDS-PAGE images were obtained by Image Lab Ver. 4.0 build 16.

Data analysis

Image processing: FIJI ver. 1.52e, FV31S-SW ver. 2.1.1.98
FCS analysis: Python ver. 3.6
AFM image processing: Falcon Viewer, IGOR Pro ver. 6.2.2.2
IDR prediction: DISOPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>)
FRAP analysis: Excel Office 365
Coalescence analysis: Origin 2019 (9.60)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for yeast experiments was determined by referring to commonly used size.
Data exclusions	No exclusion.
Replication	Quantitative FRAP experiments were repeated three times with consistent data. FCS analyses were performed for dozens of cells in order to obtain statistically reliable data. Observation of coalescence of droplets was repeated three times both in vivo and in vitro with consistent data. Observation of Ostwald ripening was repeated three times with consistent data. Phosphorylation and dephosphorylation experiments in vitro were repeated three times with consistent data. HS-AFM observation of scaffold droplets were repeated three times with consistent data. GUV experiments were repeated at least five times with consistent data. Observation of PAS localization in yeast cells was repeated three times with consistent data.
Randomization	Randomization is not relevant for this study because our work does not involve clinical trials or population studies.
Blinding	Blinding is not relevant for this study because no group allocation was performed.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Anti-FLAG M2 (F3165; Sigma; Lot#SLBN8915V)(dilution 1:1000), Anti-HA TANA2 (M180-3S; MBL; Lot#003)(dilution 1:10000), Anti-Mouse IgG (Fab specific)–Peroxidase antibody, polyclonal (A9917; Sigma; Lot#081M4762)(dilution 1:20000), Anti-Rabbit IgG (whole molecule)–Peroxidase antibody, polyclonal (A6154; Sigma; Lot# SLBK2462V)(dilution 1:20000), Anti-Atg1-T226-P (produced in this study)(dilution 1:2000), Anti-Atg13-S428/9-P (Nat Struct Mol Biol 21, 513, 2014)(dilution 1:2000)
Validation	Validation for anti-FLAG M2 was performed by detecting a single band of protein on a western blot from an E. coli crude cell lysate: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/f3165dat.pdf ; Validation for anti-HA was performed by detecting a band of N-terminal HA fusion proteins in cell extracts: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/2/h3663dat.pdf ; anti-mouse IgG: https://www.sigmaaldrich.com/content/dam/

sigma-aldrich/docs/Sigma/Datasheet/6/a9917dat.pdf; anti-rabbit IgG: <https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/6/a0545dat.pdf>; validation data for Anti-Atg13-S428/9-P are provided in Nat Struct Mol Biol 21, 513, 2014; validation data for anti-Atg1-pThr226 are provided in extended data Fig. 2b

Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer

<https://doi.org/10.1038/s41586-020-1930-8>

Received: 26 July 2018

Accepted: 25 November 2019

Published online: 22 January 2020

Emily J. Faivre¹, Keith F. McDaniel¹, Daniel H. Albert¹, Srinivasa R. Mantena², Joshua P. Plotnik¹, Denise Wilcox¹, Lu Zhang¹, Mai H. Bui¹, George S. Sheppard¹, Le Wang¹, Vasudha Sehgal¹, Xiaoyu Lin¹, Xiaoli Huang¹, Xin Lu¹, Tamar Uziel¹, Paul Hessler¹, Lloyd T. Lam¹, Richard J. Bellin¹, Gaurav Mehta¹, Steve Fidanze¹, John K. Pratt¹, Dachun Liu¹, Lisa A. Hasvold¹, Chaohong Sun¹, Sanjay C. Panchal¹, John J. Nicolette², Stacey L. Fossey², Chang H. Park¹, Kenton Longenecker¹, Lance Bigelow¹, Maricel Torrent¹, Saul H. Rosenberg¹, Warren M. Kati¹ & Yu Shen^{1*}

Proteins of the bromodomain and extra-terminal (BET) domain family are epigenetic readers that bind acetylated histones through their bromodomains to regulate gene transcription. Dual-bromodomain BET inhibitors (DbBi) that bind with similar affinities to the first (BD1) and second (BD2) bromodomains of BRD2, BRD3, BRD4 and BRDt have displayed modest clinical activity in monotherapy cancer trials. A reduced number of thrombocytes in the blood (thrombocytopenia) as well as symptoms of gastrointestinal toxicity are dose-limiting adverse events for some types of DbBi^{1–5}. Given that similar haematological and gastrointestinal defects were observed after genetic silencing of *Brd4* in mice⁶, the platelet and gastrointestinal toxicities may represent on-target activities associated with BET inhibition. The two individual bromodomains in BET family proteins may have distinct functions^{7–9} and different cellular phenotypes after pharmacological inhibition of one or both bromodomains have been reported^{10,11}, suggesting that selectively targeting one of the bromodomains may result in a different efficacy and tolerability profile compared with DbBi. Available compounds that are selective to individual domains lack sufficient potency and the pharmacokinetics properties that are required for in vivo efficacy and tolerability assessment^{10–13}. Here we carried out a medicinal chemistry campaign that led to the discovery of ABBV-744, a highly potent and selective inhibitor of the BD2 domain of BET family proteins with drug-like properties. In contrast to the broad range of cell growth inhibition induced by DbBi, the antiproliferative activity of ABBV-744 was largely, but not exclusively, restricted to cell lines of acute myeloid leukaemia and prostate cancer that expressed the full-length androgen receptor (AR). ABBV-744 retained robust activity in prostate cancer xenografts, and showed fewer platelet and gastrointestinal toxicities than the DbBi ABBV-075¹⁴. Analyses of RNA expression and chromatin immunoprecipitation followed by sequencing revealed that ABBV-744 displaced BRD4 from AR-containing super-enhancers and inhibited AR-dependent transcription, with less impact on global transcription compared with ABBV-075. These results underscore the potential value of selectively targeting the BD2 domain of BET family proteins for cancer therapy.

Analysis of historical time-resolved fluorescence resonance energy transfer (TR-FRET) data from approximately 2,500 compounds from our DbBi program identified ethyl amide **1**, with a modest 17× selectivity for the BD2 of BRD4 compared with the BD1 of BRD4. Although activity against the BD2 of BRD4 for compound **1** (1.2 nM) was not improved

compared with the DbBi ABBV-075 (1.3 nM), activity against the BD1 of BRD4 was reduced compared with ABBV-075 (22 nM for **1**, 2.8 nM for ABBV-075). Replacement of the 2,4-difluorophenyl moiety of **1** with a 2,6-dimethylphenyl ether further impaired BD1 activity, resulting in the 110× BD2-selective pyrrolopyridone **2** (124 nM and 1.1 nM for BD1 and

¹Oncology Discovery, AbbVie, North Chicago, IL, USA. ²Preclinical Safety, Development Sciences, AbbVie, North Chicago, IL, USA. *e-mail: yu.shen@abbvie.com

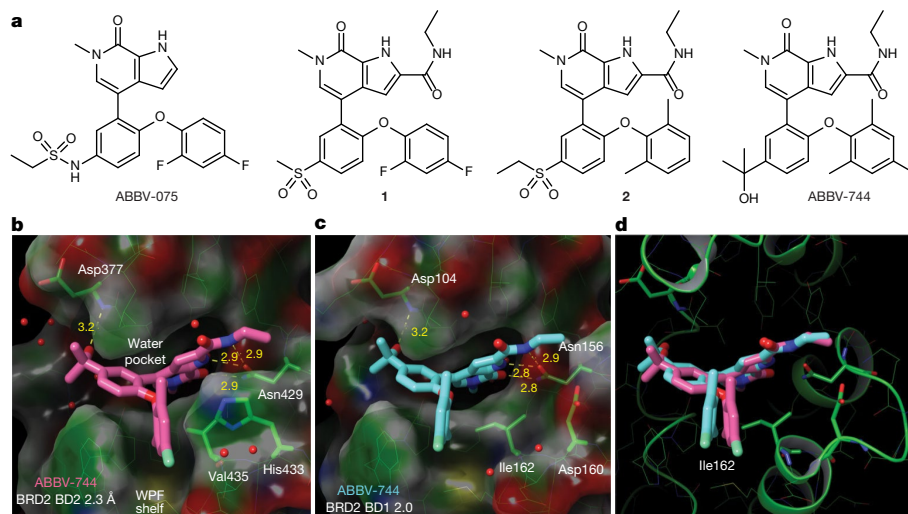


Fig. 1 | ABBV-744 is a potent and highly selective inhibitor of the BD2 domain of BET family proteins. **a**, Chemical structure of indicated compounds. **b**, Co-crystal structure of ABBV-744 (pink) in complex with BD2 of BRD2.

c, Co-crystal structure of ABBV-744 (blue) in complex with BD1 of BRD2. **d**, Overlay of the co-crystal structure of ABBV-744 in complex with BD2 of BRD2 (pink) and with BD1 of BRD2 (blue), displayed on the BRD2 BD1 protein (green).

BD2 of BRD4, respectively). Continued optimization of BD2 selectivity, metabolic stability and physical properties afforded a tertiary alcohol on the central phenyl ring in place of the ethyl sulfonamide of ABBV-075, substantially affecting the activity against BD1 of BRD4 (520 nM for ABBV-744). Addition of a fluorine atom to the phenyl ether resulted in improved pharmacokinetic properties, leading to the discovery of ABBV-744 (Fig. 1a).

ABBV-744 potently inhibited the BD2 domain of BET family proteins with more than 290× selectivity relative to the BD1 domains of BRD2, BRD3 and BRD4, and more than 95× selectivity compared with BD1 of BRD4 using TR-FRET, displayed K_d values of 3,300 nM and 2.1 nM in surface plasmon resonance experiments and half-maximum inhibitory concentrations (IC_{50}) of 20,700 nM and 27.5 nM using NanoBRET assays for BD1 and BD2 of BRD4, respectively (Extended Data Fig. 1a, b). ABBV-744 also lacked significant activity against 75 kinases and 22 bromodomain-containing proteins that represent diverse branches of the kinome and bromodome (Extended Data Fig. 1c and Supplementary Tables 1, 2). ABBV-744 is primarily metabolized by CYP3A4 and shows oral bioavailability, enabling in vivo efficacy and tolerability studies (Extended Data Fig. 1d, e).

The crystal structures of ABBV-744 complexed with both the BD2 and BD1 of BRD2 established the binding mode of ABBV-744 that underlies its BD2 selectivity (Fig. 1b–d and Extended Data Table 1). ABBV-744 maintains all of the important interactions found for canonical DbBi^{14–16}, including binding of the pyrrolopyridone with the conserved Asn156 residue, placement of the *N*-methyl moiety in the amphipathic water pocket, and positioning of an aryl ring in the WPF shelf in both BD2 and BD1 (Fig. 1b, c). The ethyl amide moiety of ABBV-744 exploits the Asp (BD1) and His433 (BD2) divergence conserved across all bromodomain BET family members by burying the amide in a channel formed by the His433, Tyr386 and Pro430 residues of BD2 (Fig. 1b), a binding interaction that is not available in BD1 (Fig. 1c). The 2,6-dimethylphenyl ether moiety of ABBV-744 targets the subtle size distinction of the Ile162 (BD1) and Val435 (BD2) sequence differences. Thus, incorporation of a dimethylphenyl ether moiety forces an aryl methyl group to be buried in the rigid base of the WPF shelf. The smaller BD2 Val435 residue can accommodate this added methyl group interaction without disruption of binding, and therefore binding potency is maintained. For the BD1 protein, however, interaction of this aryl methyl group with the larger Ile162 residue forces the inhibitor to shift slightly away from the Ile moiety, causing a subtle change in the placement of both the

aryl group and the hydroxy group of the tertiary alcohol, leading to a less-optimal binding interaction (overlay in Fig. 1d) and a decrease in the potency of ABBV-744 with BD1.

We tested ABBV-744 in 59 cancer cell lines that are sensitive to DbBi^{17–22} and found that ABBV-744 retained robust antiproliferative activity (IC_{50} < 100 nM) mostly—but not exclusively—in acute myeloid leukaemia cells and a subset of prostate cancer cells that expressed the full-length AR, but not those expressing AR-V7 or that were AR-negative (Fig. 2a, Extended Data Table 2 and Supplementary Fig. 1). Similar to ABBV-075 and the AR antagonist enzalutamide, ABBV-744 induced cell cycle arrest in G1 followed by senescence in LNCaP cells (Fig. 2b). Narrow antiproliferative activity was also observed for a structurally distinct compound, A-083, across 240 cancer cell lines (Extended Data Fig. 2a–c and Supplementary Table 3). BD2 inhibitor compounds 74, 75 and RVX-208^{11,12} displayed a similar albeit weaker antiproliferative trend, probably owing to their moderate selectivity and weaker binding affinity (Extended Data Fig. 2d). Relative to ABBV-075, ABBV-744 demonstrated limited potency in viability assays of megakaryocyte colony forming units (Mk-CFU) in mice and IEC-6 cells, which are potential surrogate assays for platelet production and proliferation of normal intestinal epithelium, respectively²³ (Extended Data Fig. 2e).

In ABBV-744-sensitive LNCaP cells, ABBV-744 elicited far fewer gene expression changes than ABBV-075 at doses at which BD2 of BRD4 was similarly inhibited or the reported doses of the DbBi JQ1 and iBET^{20,24} (Fig. 2c and Extended Data Fig. 3a). For example, at a BD2-selective concentration (48 nM), ABBV-744 downregulated *ACPP* (also known as *ACP3*) and *MYC* but did not affect the ABBV-075-responsive genes *HEXIM1*, *SPDEF* and *ZG16B* (Fig. 2d and Extended Data Fig. 3b). The BD2-dependent genes *KLK2* and *MYC* were also partially inhibited by a potent and selective BD1 inhibitor described in the patent literature (BDIi)¹³, suggesting that these BD2-dependent genes were in part also dependent on BD1, and combined blockade of both domains mimicked the activity of ABBV-075 (Extended Data Fig. 3c). When used at a high concentration (6 μ M), ABBV-744 probably engaged both BD1 and BD2, thus recapitulating the activities of ABBV-075 against all genes (Fig. 2d). Notably, this small set of 241 ABBV-744-regulated genes is highly enriched in dihydrotestosterone (DHT)-responsive genes (Fig. 2e and Supplementary Table 4). Gene set enrichment analysis also revealed common regulation of AR, MYC and E2F1 hallmarks by ABBV-744, enzalutamide and ABBV-075, similar to reports for JQ1 and iBET^{20,24} (Extended Data Fig. 3d). Although both ABBV-744 and ABBV-075

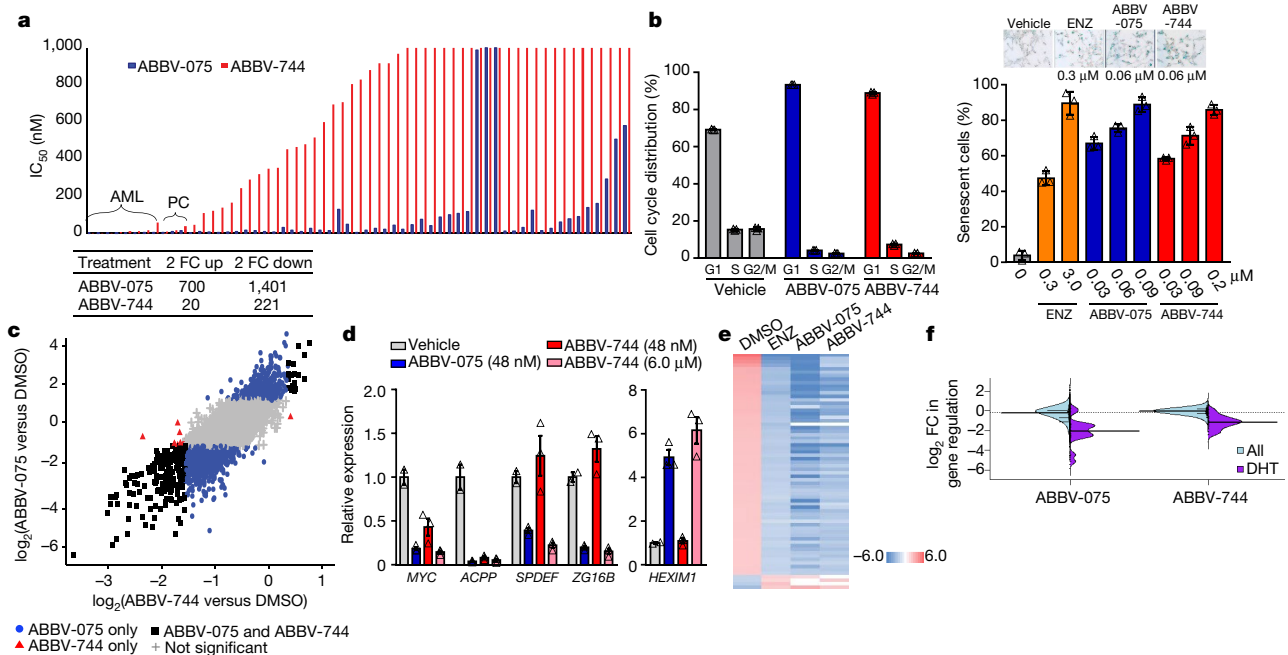


Fig. 2 | ABBV-744 exhibits potent antiproliferative activity against AR-positive prostate cancer cells and inhibits AR-dependent transcription. a, The antiproliferative IC_{50} values across cancer cell lines after treatment with ABBV-075 or ABBV-744 for 5 days. **b**, ABBV-744 induced cell cycle arrest (left, 72 h 60 nM ABBV-075 or 90 nM ABBV-744; concentrations that elicited similar degrees of inhibition of BD2 of BRD4) and senescence (right, 12 days). ENZ, enzalutamide. Data are mean \pm s.d. ($n = 3$ biologically independent samples) and are representative of $n = 3$ independent experiments. Representative images of β -galactosidase staining of cells at 100 \times magnification are shown in the top right. **c**, Number of significantly regulated genes (fold change in expression > 2 -fold, $P < 0.01$, $n = 2$, statistical analysis by DESeq2 algorithm) and scatter plot of \log_2 -transformed fold change in expression after 24 h treatment compared with DHT stimulation alone in phenol red-free, charcoal stripped serum (vehicle, 5 nM DHT, 5 nM DHT and 60 nM ABBV-075, or 5 nM DHT and 90 nM ABBV-744). Genes significantly regulated by both ABBV-075 and ABBV-744 or by individual compounds were labelled as ABBV-075 and

ABBV-744, ABBV-075 only, or ABBV-744 only. **d**, Expression of BD2-sensitive and -insensitive genes quantified using the branched DNA (bdNA) assay after treatment for 24 h with ABBV-075 or ABBV-744 in the presence of 5 nM DHT. Data are mean \pm s.d. ($n = 3$ biological replicates) and are representative of $n = 2$ independent experiments. **e**, Heat map of DHT-induced gene expression alterations (DHT signature, fold change in expression of > 2 , $P < 0.01$ for DHT versus vehicle, $n = 2$) and the response of these DHT signature genes to treatment with enzalutamide, ABBV-075 or ABBV-744 in DHT-stimulated cells. **f**, Genes significantly ($q < 0.01$) regulated by ABBV-075 or ABBV-744 in DHT-stimulated cells were classified as DHT-regulated genes (overlapping with the DHT signature) or non-DHT regulated genes (outside of the DHT signature). The distribution of \log_2 -transformed fold changes in expression is shown as a split violin plot. The long solid line represents the mean fold change. The small lines represent individual data points. The dotted line represents the overall average. Statistical significance between all and DHT was determined by two-tailed unpaired Student's t -test, P values were calculated by DESeq2.

prominently downregulated the DHT signature, ABBV-075 induced a broader distribution of expression alterations than ABBV-744 and affected hallmarks that were not affected by ABBV-744 (Fig. 2f and Extended Data Fig. 3d). Collectively, these results suggest that ABBV-744 significantly inhibited AR-dependent transcription in LNCaP cells while having a lower impact on global transcription than ABBV-075.

DbBi has been shown to downregulate AR protein expression in some but not all experimental settings, probably owing to subtle differences in cell lines and exact experimental conditions. In our hands, neither ABBV-075 nor ABBV-744 reduced AR protein levels in LNCaP cells (Extended Data Fig. 3b). Given the lack of a direct effect on the AR protein, genome-wide AR and BRD4 occupancy was determined to understand the sensitivity of AR-dependent transcription to ABBV-744 in LNCaP cells. ABBV-075 but not ABBV-744 caused AR peak loss similar to JQ1 treatment²⁰ (Extended Data Fig. 4a). Dependency profiles from the DepMap portal (<https://depmap.org/portal/>) indicated that prostate cancer cell lines are significantly more dependent on BRD4 than BRD2 or BRD3, and higher BRD4 dependency is associated with higher sensitivity to ABBV-744 (Extended Data Fig. 4b), collectively suggesting that BRD4 may be the primary BET family driver of prostate cancer cell line viability and an important target of ABBV-744. ABBV-744 displayed a globally weaker but otherwise similar pattern of BRD4 peak displacement relative to ABBV-075 and JQ1, and preferentially downregulated genes associated with super-enhancers similar to DbBi^{20,25,26}

(Fig. 3a–c and Extended Data Fig. 4c). A subset of BRD4 peaks overlapped with AR peaks, and notably 43% of the BRD4/AR co-occupied sites were in super-enhancers (Extended Data Fig. 4d). Interestingly, BRD4 was highly bound at AR-occupied super-enhancers relative to non-AR super-enhancers, and ABBV-744 and ABBV-075 both effectively displaced BRD4 from the AR-containing super-enhancers, suggesting an increased dependence of BRD4–AR co-occupied super-enhancers on BD2 (Fig. 3d). Further integrating the BRD4-binding profile with gene regulation by AR, ingenuity pathway analysis and motif analysis revealed enrichment of DHT pathway and androgen-response elements within super-enhancers from which BRD4 was displaced by both ABBV-744 and ABBV-075 (Extended Data Fig. 5a, b). For example, ABBV-744 displaced BRD4 from BRD4–AR co-occupied super-enhancers that are closely associated with AR-dependent genes and inhibited KLK2 expression (Fig. 3e–g and Extended Data Fig. 3b). Similarly, ABBV-744 significantly affected BRD4 occupancy on super-enhancers associated with BD2-sensitive *ACPP* but not BD2-insensitive *ZG16B* (Extended Data Fig. 5c).

To understand the sensitivity of BRD4–AR co-occupied super-enhancers to ABBV-744, we tested BD2 dependency of the reported BRD4–AR interaction²⁰. A small but reproducibly detectable fraction of BRD4 was found in complex with AR. This DHT and acetylation-dependent interaction was disrupted by ABBV-744 and ABBV-075. By contrast, the reported interactions of BRD4 with CDK9, GATA2 or CDK9/cyclin T1 with HEXIM1 were not BD2 dependent^{7,27,28} (Extended Data Fig. 6a–c).

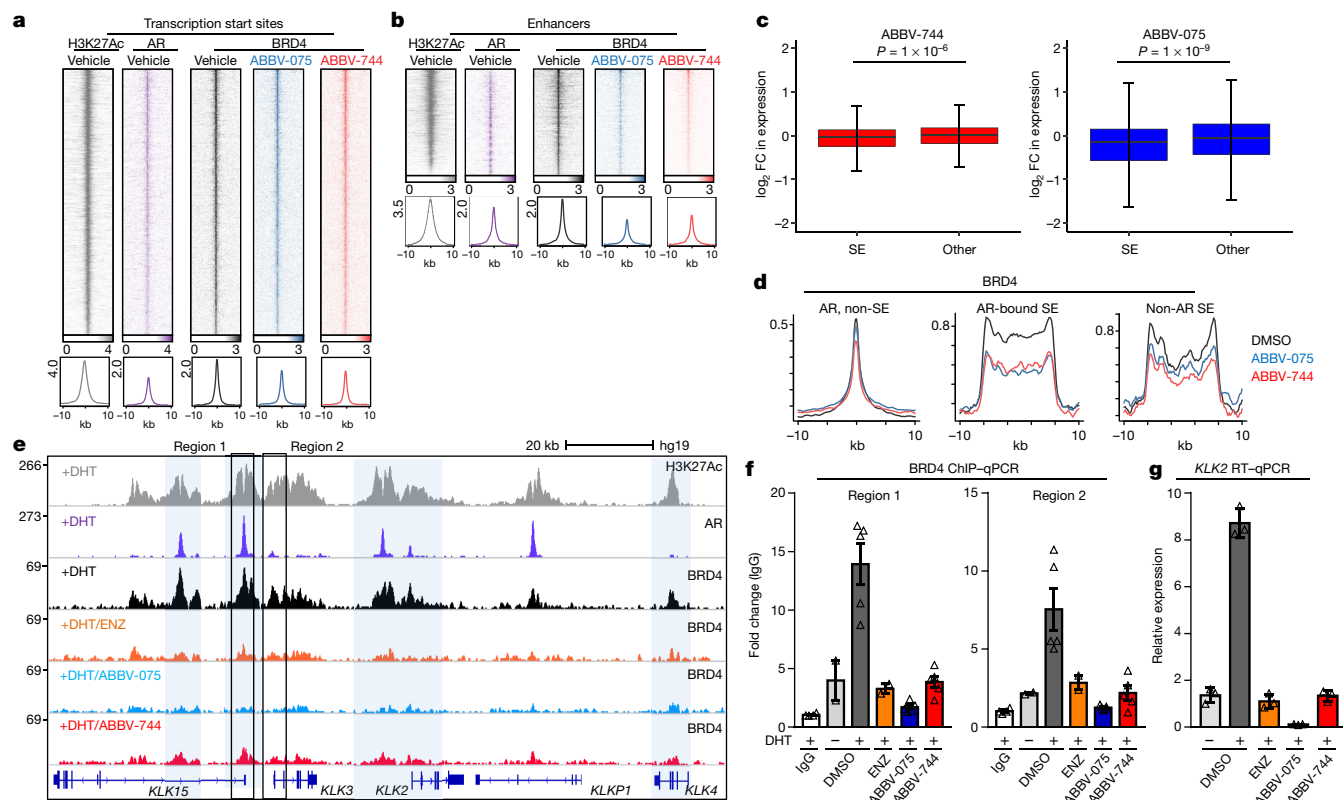


Fig. 3 | ABBV-744 displaces BRD4 from AR-containing super-enhancers. LNCaP cells were incubated with 5 nM DHT and vehicle (DMSO), 60 nM ABBV-075 or 90 nM ABBV-744 for 6 h, and cells were collected for ChIP-seq to determine H3K27Ac, AR and BRD4 chromatin association. **a, b**, Rank-ordered heat maps of H3K27Ac, AR and BRD4 peaks at transcription start sites or enhancers after the indicated treatment. Rows are ordered according to the vehicle-treated BRD4 maximum for each region and centred ± 10 kb of the BRD4 peak after treatment with vehicle. Colour scales depict reads per million (RPM) intensities. Bottom profile plots display \log_2 -transformed fold change in RPM/bp compared with control. BRD4 ChIP experiments were normalized to spike-in controls. **c**, Quantification of \log_2 -transformed fold change in expression after ABBV-075 or ABBV-744 treatment for genes associated with super-enhancers (SE) or non-super-enhancers (other). For all box plots, centre line indicates the median; box limits are the first and third quartiles; whiskers

range from the first quartile minus $1.5 \times$ the interquartile range to the third quartile plus $1.5 \times$ the interquartile range. Unpaired two-tailed Student's *t*-test was used to determine significance for super-enhancers versus other; $n = 2$. **d**, BRD4 profile plots at AR-bound regions that are not located in super-enhancers (AR, non-SE), AR-bound super-enhancers (AR, SE), or super-enhancers without AR binding (non-AR SE). **e**, Gene track of H3K27Ac, AR, and BRD4 ChIP-seq signals for the indicated treatment conditions at a super-enhancer that is associated with several AR-dependent genes. **f**, LNCaP cells that underwent the indicated treatments for 24 h were collected for ChIP-qPCR to determine the binding of BRD4 to the indicated regions in the gene track. **g**, *KLK2* expression in LNCaP cells that underwent the indicated treatments for 24 h was determined by qPCR. **f, g**, Data are mean \pm s.d. ($n = 3$ biologically independent samples) and are representative of $n > 2$ independent experiments.

Notably, AR acetylation at the $K_{630}LKK_{633}$ motif that resembles BET bromodomain-binding sites in histones has been shown to be important for AR activity²⁹. Considering that the N-terminal domain of AR has been shown to bind to BD1 directly²⁰, we speculated that acetylated AR may interact cooperatively with both BD1 and BD2 of BRD4 at AR-BRD4 co-occupied super-enhancers to regulate a subset of AR-dependent genes that are therefore sensitive to BD2 inhibition (Extended Data Fig. 6d, e). In ABBV-744-resistant 22RV1 cells, in which AR-dependent transcription is driven by AR-V7 (which lacks the $K_{630}LKK_{633}$ motif³⁰), ABBV-744 failed to inhibit the AR gene signature, induced limited BRD4 displacement from super-enhancers, and produced weak effects on proliferation and senescence, collectively supporting the putative interaction of acetylated AR with BD2 to induce sensitivity to ABBV-744 (Extended Data Fig. 7a–f). More mechanistic studies will be required to confirm this hypothesis.

The drug-like properties of ABBV-744 enabled the investigation of its antitumour efficacy and tolerability. In a mouse xenograft model using LNCaP cells, treatment with 4.7 mg kg^{-1} ABBV-744 (1/16 of the maximum tolerated dose (MTD)) caused a delay in tumour growth that was equivalent to ABBV-075 treatment at the MTD dose of 1 mg kg^{-1} (Fig. 4a). Comparing efficacious exposure levels of ABBV-744 in LNCaP

tumour-bearing mice (4.7 mg kg^{-1} ; area under the curve, $1.1 \mu\text{g h ml}^{-1}$) and MTD (75 mg kg^{-1} ; area under the curve, $13.1 \mu\text{g h ml}^{-1}$) demonstrated that ABBV-744 was able to produce significant antitumour activity at 1/12 of the highest tolerable exposure of ABBV-744 (Extended Data Fig. 8a). The activity exhibited by ABBV-744 at 1/16 of the MTD of ABBV-744 was superior to the activities achieved using JQ1 and iBET at their respective MTDs or, in the case of RVX-208, at the highest feasible dose in this model (Extended Data Fig. 8b, c). Similarly, ABBV-744 at 1/16 MTD also displayed equivalent or better antitumour activity compared with ABBV-075 at MTD in the enzalutamide-resistant MDA-PCa-2b xenograft model (Fig. 4b). As a control, lowering the dose of ABBV-075 to 1/2 of the MTD resulted in a significant reduction in antitumour activity to 42% tumour growth inhibition in the LNCaP xenograft model. Even in the xenograft model using OPM2 cells, one of the most sensitive models to DbBi, ABBV-075 at 1/4 of the MTD of ABBV-075 (0.25 mg kg^{-1}) had only marginal antitumour efficacy (Extended Data Fig. 8d, e).

In toxicity studies in rats, ABBV-075 at 3 mg kg^{-1} ($3 \times$ the efficacious exposure in the LNCaP mouse xenograft model), caused a 59% reduction in platelets, a decrease in Alcian blue staining of the mucosa and the loss of goblet cells. By contrast, ABBV-744 at 30 mg kg^{-1} ($25 \times$ the efficacious exposure) triggered a reduction in platelets of only 20%, and at

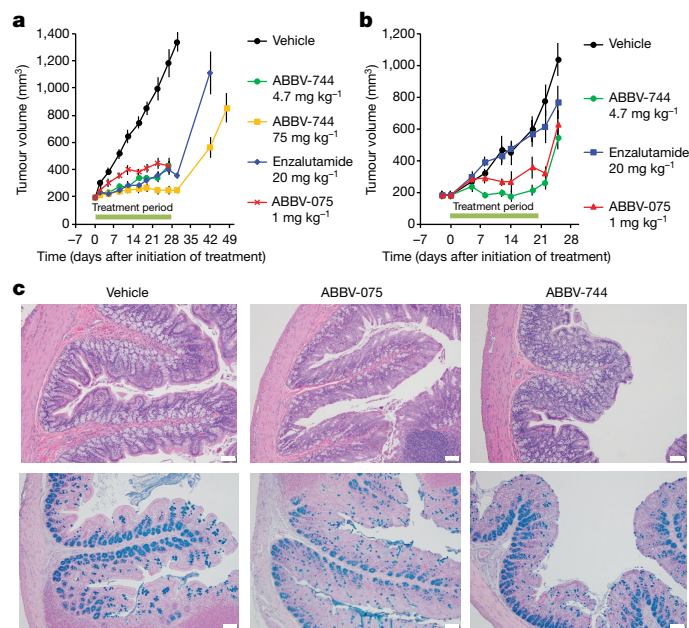


Fig. 4 | ABBV-744 maintains DbBi-like activity in AR positive prostate cancer xenografts while displaying an improved tolerability profile. a, b, Mice bearing LNCaP (a) or MDA-PCa-2b tumours (b) were treated daily with enzalutamide, ABBV-075 or ABBV-744 at the indicated amounts using oral gavage throughout the indicated treatment period. Data are mean \pm s.e.m. ($n = 9$ mice per group in a) and 7 mice per group in b). Mice treated with 4.7 mg kg⁻¹ ABBV-744 or 1 mg kg⁻¹ ABBV-075 were euthanized on day 28 to conduct ancillary studies. **c,** Sprague-Dawley rats ($n = 3$ animals per group) were treated daily with vehicle, 3 mg kg⁻¹ ABBV-075 or 60 mg kg⁻¹ ABBV-744 for 14 days. Histopathology assessment was carried out using large-intestinal sections after necropsy. Alcian blue staining was used to characterize goblet cells. Representative images of haematoxylin and eosin staining (top) and alcian blue staining (bottom) are shown. Efficacious exposure levels of ABBV-075 (1 mg kg⁻¹) and ABBV-744 (4.7 mg kg⁻¹) in mice and exposure levels associated with the indicated doses of each compound in rats were determined in separate animals used for pharmacokinetic studies ($n = 3$ animals).

60 mg kg⁻¹ (47 \times the efficacious exposure) did not cause loss of goblet cells or other gross intestinal defects (Fig. 4c and Extended Data Fig. 8a). Similarly, 2.5 mg kg⁻¹ ABBV-075 caused germ cell degeneration in the testes, whereas no microscopic changes in the testes were observed with 25 mg kg⁻¹ ABBV-744. These efficacy and tolerability results collectively suggest that selectively targeting BD2 can induce antitumour activity in some cancer settings while mitigating key tolerability issues of DbBi. These findings support the advancement of ABBV-744 for clinical evaluation (ClinicalTrials.gov identifier NCT03360006) and call for further investigation of BD2-dependent transcription programs to reveal additional therapeutic opportunities.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1930-8>.

- Amorim, S. et al. Bromodomain inhibitor OTX015 in patients with lymphoma or multiple myeloma: a dose-escalation, open-label, pharmacokinetic, phase 1 study. *Lancet Haematol.* **3**, e196–e204 (2016).
- Stathis, A. et al. Clinical response of carcinomas harboring the BRD4-NUT oncoprotein to the targeted bromodomain inhibitor OTX015/MK-8628. *Cancer Discov.* **6**, 492–500 (2016).
- Abramson, J. S. et al. BET inhibitor CPI-0610 is well tolerated and induces responses in diffuse large B-cell lymphoma and follicular lymphoma: preliminary analysis of an ongoing phase 1 study. *Blood* **126**, 1491 (2015).
- O'Dwyer, P. J. et al. Abstract CT014: GSK525762, a selective bromodomain (BRD) and extra terminal protein (BET) inhibitor: results from part 1 of a phase I/II open-label single-agent study in patients with NUT midline carcinoma (NMC) and other cancers. *Cancer Res.* **76**, CT014 (2016).
- Piha-Paul, S. A. et al. Results of the first-in-human study of ABBV-075 (mivebresib), a pan-inhibitor of bromodomain (BD) and extra terminal (BET) proteins, in patients (pts) with relapsed/refractory (R/R) solid tumors. *J. Clin. Oncol.* **36**, 2510 (2018).
- Bolden, J. E. et al. Inducible in vivo silencing of Brd4 identifies potential toxicities of sustained BET protein inhibition. *Cell Rep.* **8**, 1919–1929 (2014).
- Gamsjaeger, R. et al. Structural basis and specificity of acetylated transcription factor GATA1 recognition by BET family bromodomain protein Brd3. *Mol. Cell. Biol.* **31**, 2632–2640 (2011).
- Shi, J. et al. Disrupting the interaction of BRD4 with diacetylated Twist suppresses tumorigenesis in basal-like breast cancer. *Cancer Cell* **25**, 210–225 (2014).
- Lamonica, J. M. et al. Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proc. Natl Acad. Sci. USA* **108**, E159–E168 (2011).
- Gacias, M. et al. Selective chemical modulation of gene transcription favors oligodendrocyte lineage progression. *Chem. Biol.* **21**, 841–854 (2014).
- Picaud, S. et al. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl Acad. Sci. USA* **110**, 19754–19759 (2013).
- Law, R. P. et al. Discovery of tetrahydroquinolines as bromodomain and extra-terminal domain (BET) inhibitors with selectivity for the second bromodomain. *J. Med. Chem.* **61**, 4317–4334 (2018).
- Dominique Amans, P. B. et al. Furopyridines as bromodomain inhibitors. International patent application PCT/EP2014/054796 (2014).
- McDaniel, K. F. et al. Discovery of N-(4-(2,4-difluorophenoxy)-3-(6-methyl-7-oxo-6,7-dihydro-1H-pyrrolo[2,3-c]pyridin-4-yl)phenyl)ethanesulfonamide (ABBV-075/mivebresib), a potent and orally available bromodomain and extraterminal domain (BET) family bromodomain inhibitor. *J. Med. Chem.* **60**, 8369–8384 (2017).
- Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
- Nicodeme, E. et al. Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119–1123 (2010).
- Bui, M. H. et al. Preclinical characterization of BET Family bromodomain inhibitor ABBV-075 suggests combination therapeutic strategies. *Cancer Res.* **77**, 2976–2989 (2017).
- Mertz, J. A. et al. Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc. Natl Acad. Sci. USA* **108**, 16669–16674 (2011).
- Puissant, A. et al. Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov.* **3**, 308–323 (2013).
- Asangani, I. A. et al. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **510**, 278–282 (2014).
- Lockwood, W. W., Zejnullahu, K., Bradner, J. E. & Varmus, H. Sensitivity of human lung adenocarcinoma cell lines to targeted inhibition of BET epigenetic signaling proteins. *Proc. Natl Acad. Sci. USA* **109**, 19408–19413 (2012).
- Yang, L. et al. Repression of BET activity sensitizes homologous recombination-proficient cancers to PARP inhibition. *Sci. Transl. Med.* **9**, eaat1645 (2017).
- Pessina, A. et al. Application of human CFU-Mk assay to predict potential thrombocytotoxicity of drugs. *Toxicol. In Vitro* **23**, 194–200 (2009).
- Wyce, A. et al. Inhibition of BET bromodomain proteins as a therapeutic approach in prostate cancer. *Oncotarget* **4**, 2419–2429 (2013).
- Winter, G. E. et al. BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol. Cell* **67**, 5–18 (2017).
- Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
- Yang, Z. et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19**, 535–545 (2005).
- Jang, M. K. et al. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19**, 523–534 (2005).
- Fu, M. et al. Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. *Mol. Cell. Biol.* **23**, 8563–8575 (2003).
- Hu, R. et al. Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69**, 16–22 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RNA-seq and ChIP-seq dataset can be accessed from GEO (accession numbers GSE118152, GSE118247 and GSE130269). Crystal coordinates and X-ray diffraction data were deposited in the Protein Data Bank with the accession numbers 6E6J and 6ONY.

Acknowledgements We thank Z. Zha for technical assistance with ChIP-seq data analysis. For X-ray crystallography, use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Hauptman-Woodward Medical Research Institute. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357.

Author contributions K.F.M., G.S.S., L.W., S.F., J.K.P., D.L. and L.A.H. designed and synthesized the compounds. E.J.F., D.W., M.H.B., X. Lin, X.H., P.H., L.Z. and R.J.B. performed in vitro studies including cell proliferation, gene expression and ChIP studies. J.P.P., V.S., T.U., P.H., L.T.L., X. Lu and E.J.F. analysed RNA-seq and ChIP-seq data. D.H.A. and G.M. performed in vivo efficacy studies. C.H.P., K.L., L.B. and M.T. contributed to three-dimensional structure data generation and analysis. S.C.P. and C.S. generated surface plasmon resonance-binding data. S.R.M., J.J.N. and S.L.F. carried out rat toxicology studies. E.J.F., D.H.A., S.R.M., J.J.N., S.L.F., W.M.K., K.F.M., S.H.R., L.Z., W.M.K. and Y.S. designed studies and interpreted results. E.J.F., K.F.M. and Y.S. wrote the paper.

Competing interests E.J.F., K.F.M., D.H.A., S.R.M., L.Z., M.H.B., G.S.S., L.W., J.P.P., V.S., X. Lin, X.H., X. Lu, T.U., L.T.L., R.J.B., G.M., S.F., J.K.P., D.L., L.A.H., C.S., S.C.P., J.J.N., S.L.F., K.L., L.B., M.T., S.H.R., W.M.K. and Y.S. are employees of AbbVie. C.H.P., D.W. and P.H. were employees of AbbVie at the time of the study. The design, study conduct and financial support for this research were provided by AbbVie. AbbVie participated in the interpretation of data, review and approval of the publication.

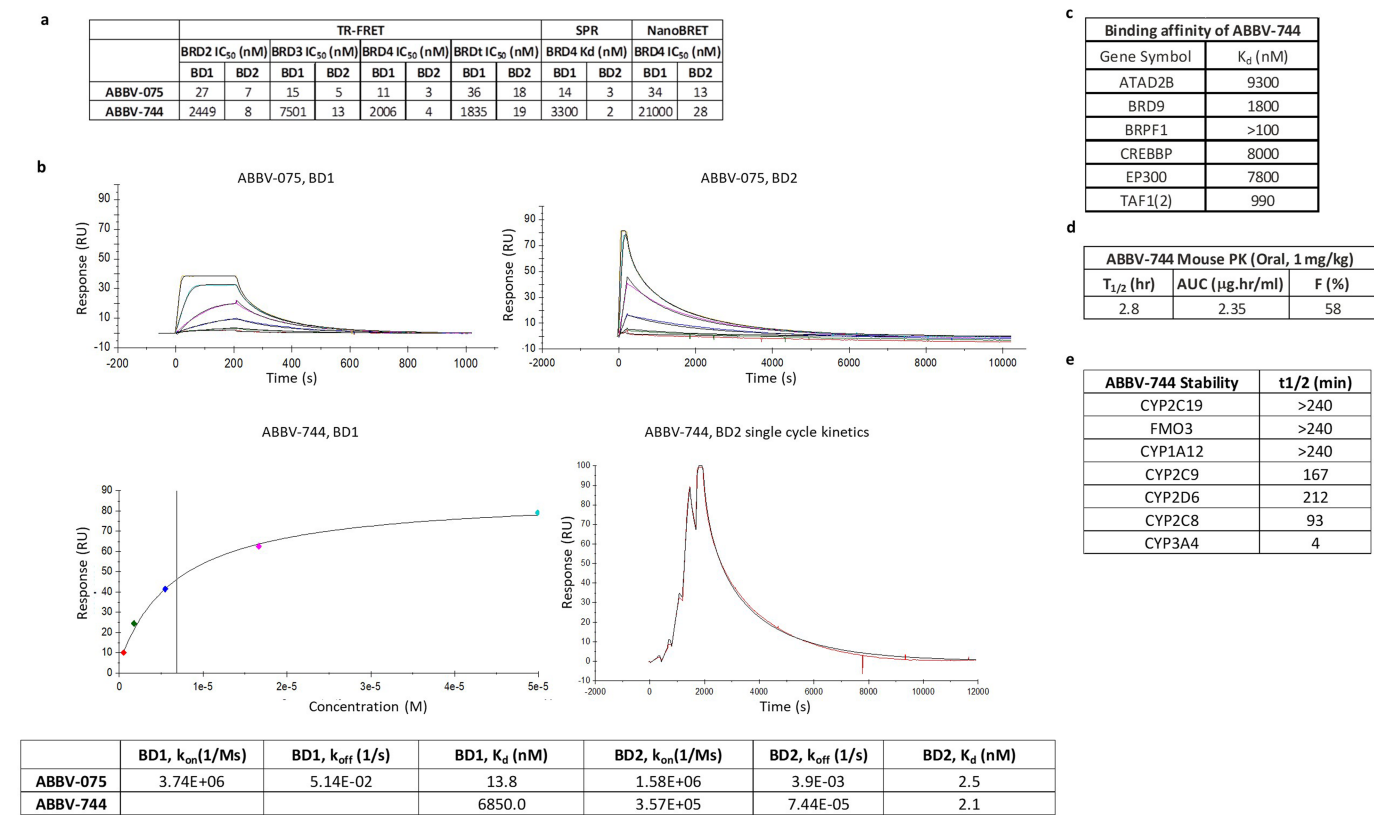
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1930-8>.

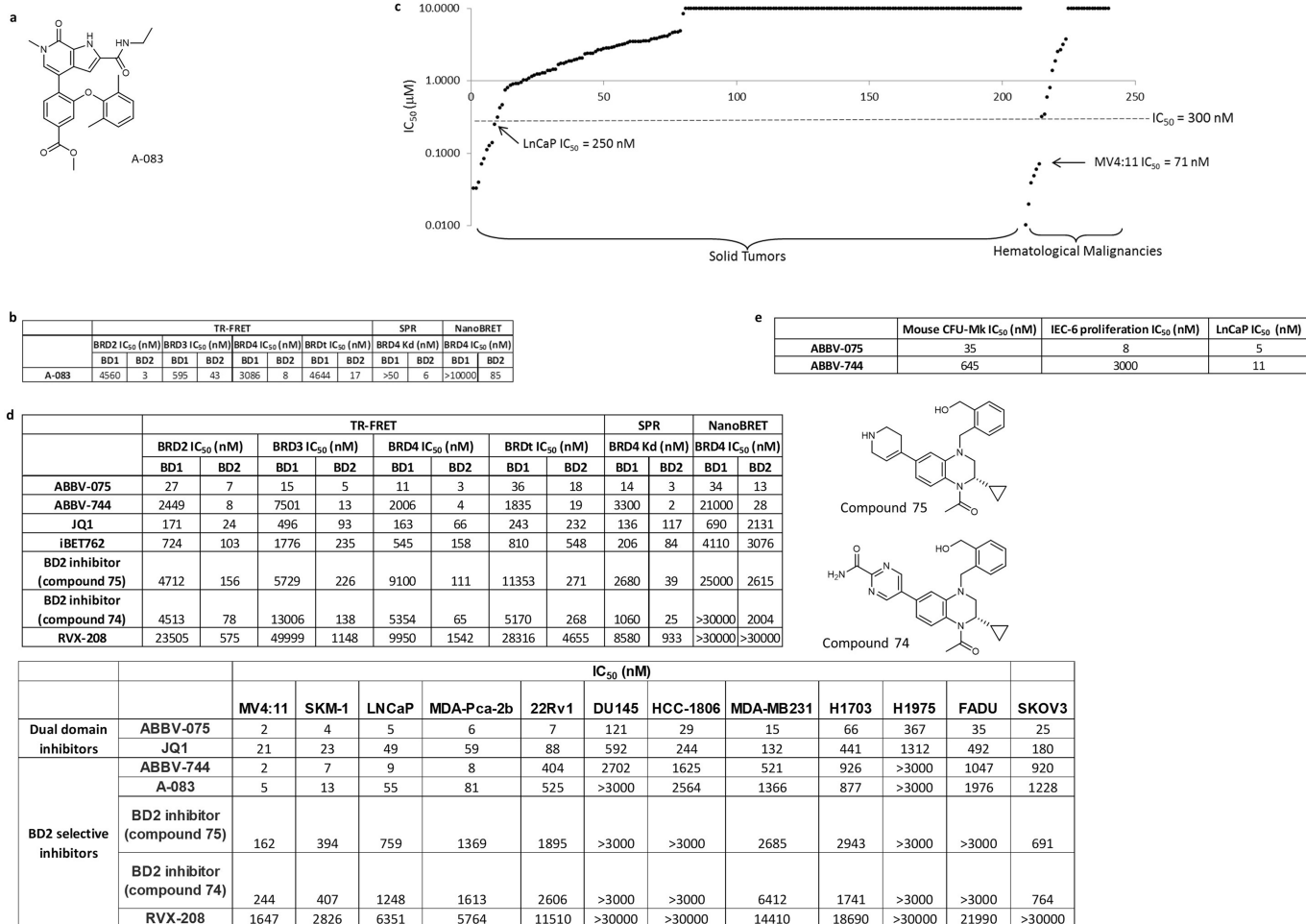
Correspondence and requests for materials should be addressed to Y.S.

Peer review information *Nature* thanks Arul Chinnaiyan, Stefan Knapp, William Pomerantz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

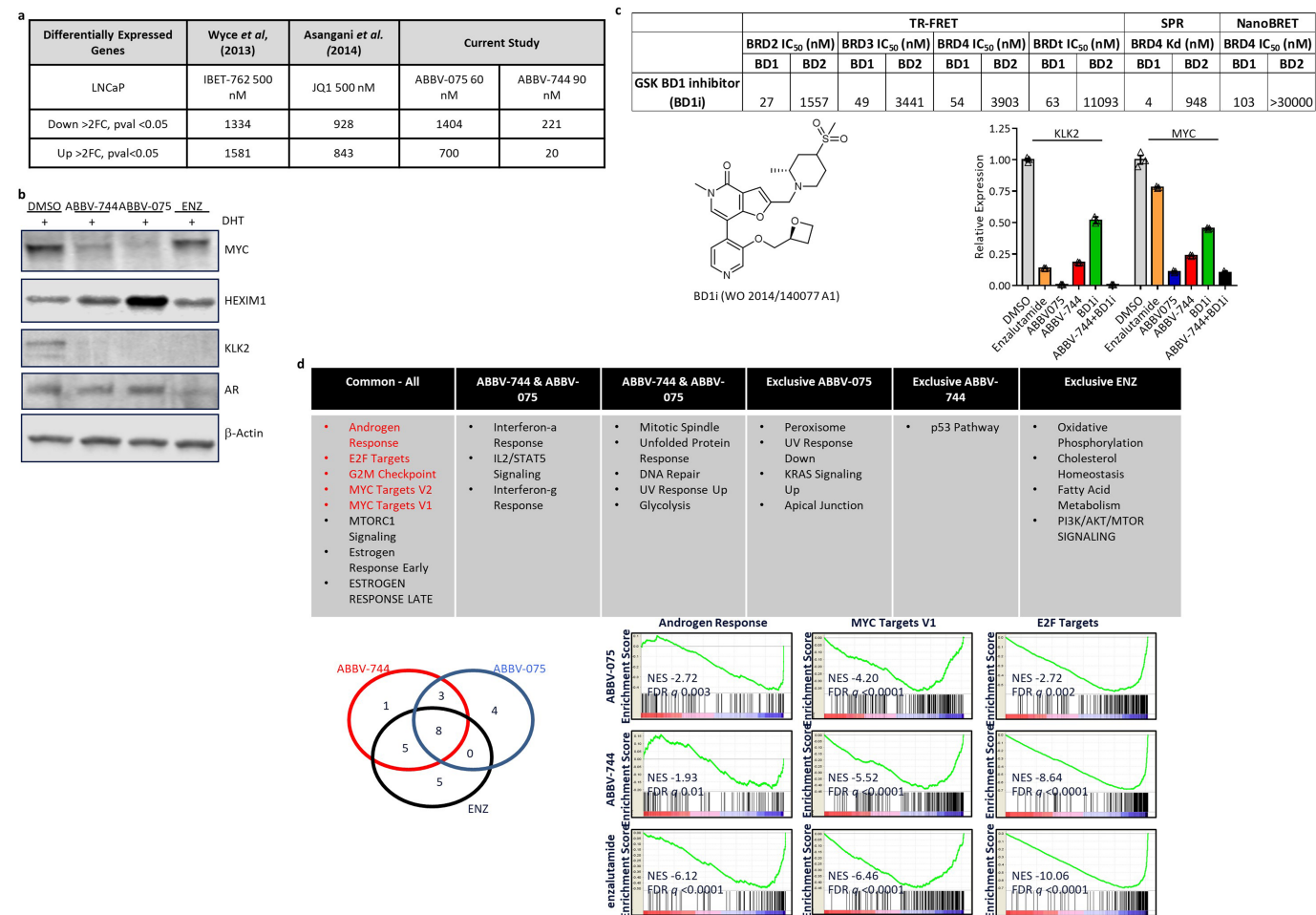


Extended Data Fig. 1 | Characterization of ABBV-744. **a**, TR-FRET, surface plasmon resonance (SPR) and NanoBRET potency and selectivity of ABBV-744. **b**, Surface plasmon resonance binding of ABBV-075 and ABBV-744 to BD1 and BD2 domains of BRD4. ABBV-075 binding curves (coloured) with fits to the 1:1 binding model (black). ABBV-744 binds to BD1 with very fast on and off kinetics, therefore a steady-state fit to equilibrium responses was used to determine Biacore affinities. Dissociation of ABBV-744 from BD2 is very slow and therefore binding was profiled using the single-cycle kinetics method. All experiments were repeated once with similar results. **c**, Binding affinities of ABBV-744 to selected bromodomains for which ABBV-744 exhibited more than 50% inhibition at 1 μM using BromoScan profiling. **d**, Pharmacokinetic parameters in mice. **e**, ABBV-744 stability after incubation with various CYP enzymes.



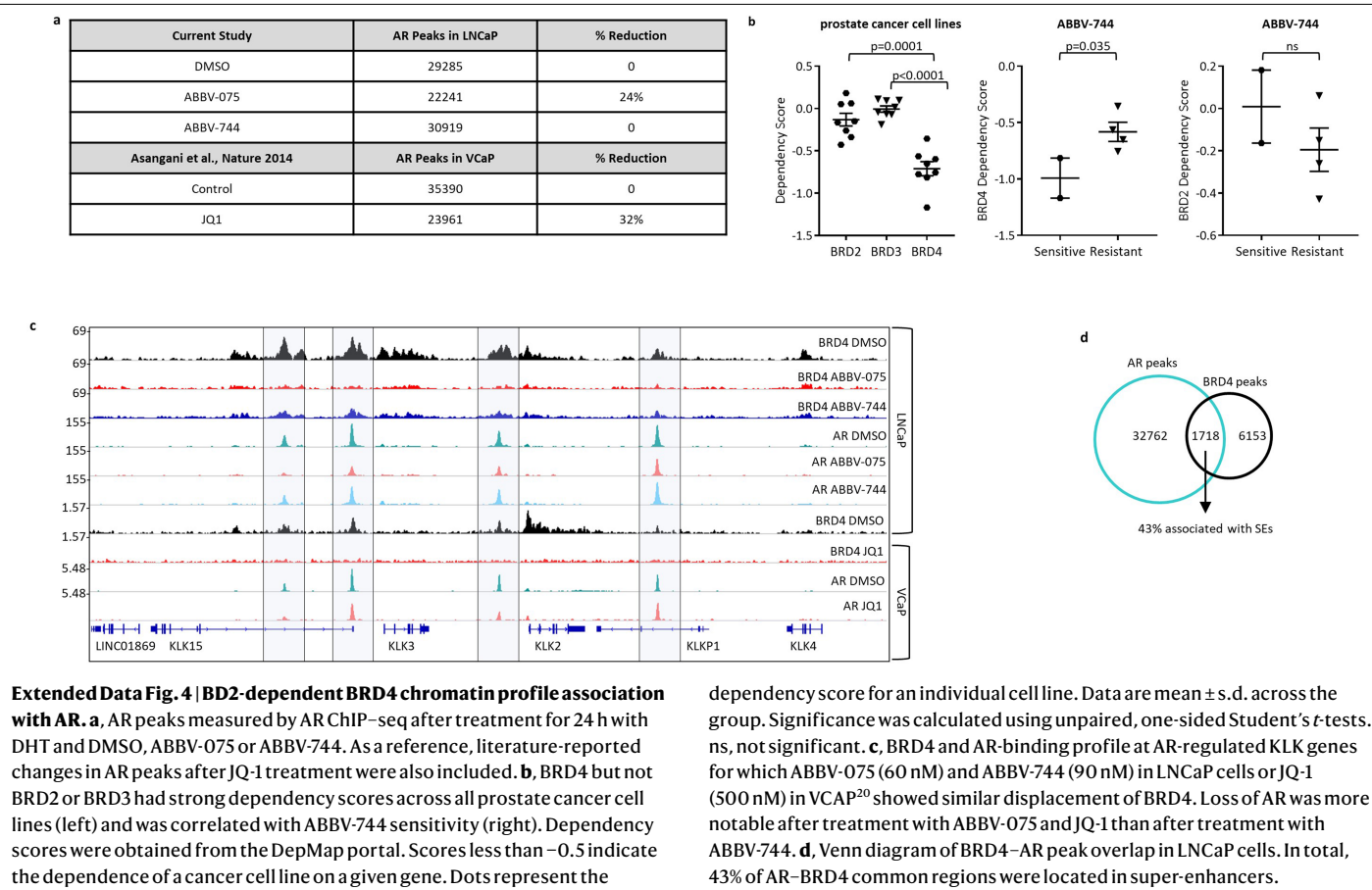
Extended Data Fig. 2 | Antiproliferative activity of structurally diverse BD2 and DbBis. a, Chemical structure of A-083. **b**, Activity of A-083 across multiple assays. **c**, Anti-proliferative activity of A-083 across the OncoPanel of Europhorin, which consist of 240 cancer cell lines across a broad spectrum of cancer indications. **d**, Characterization and antiproliferative activities of

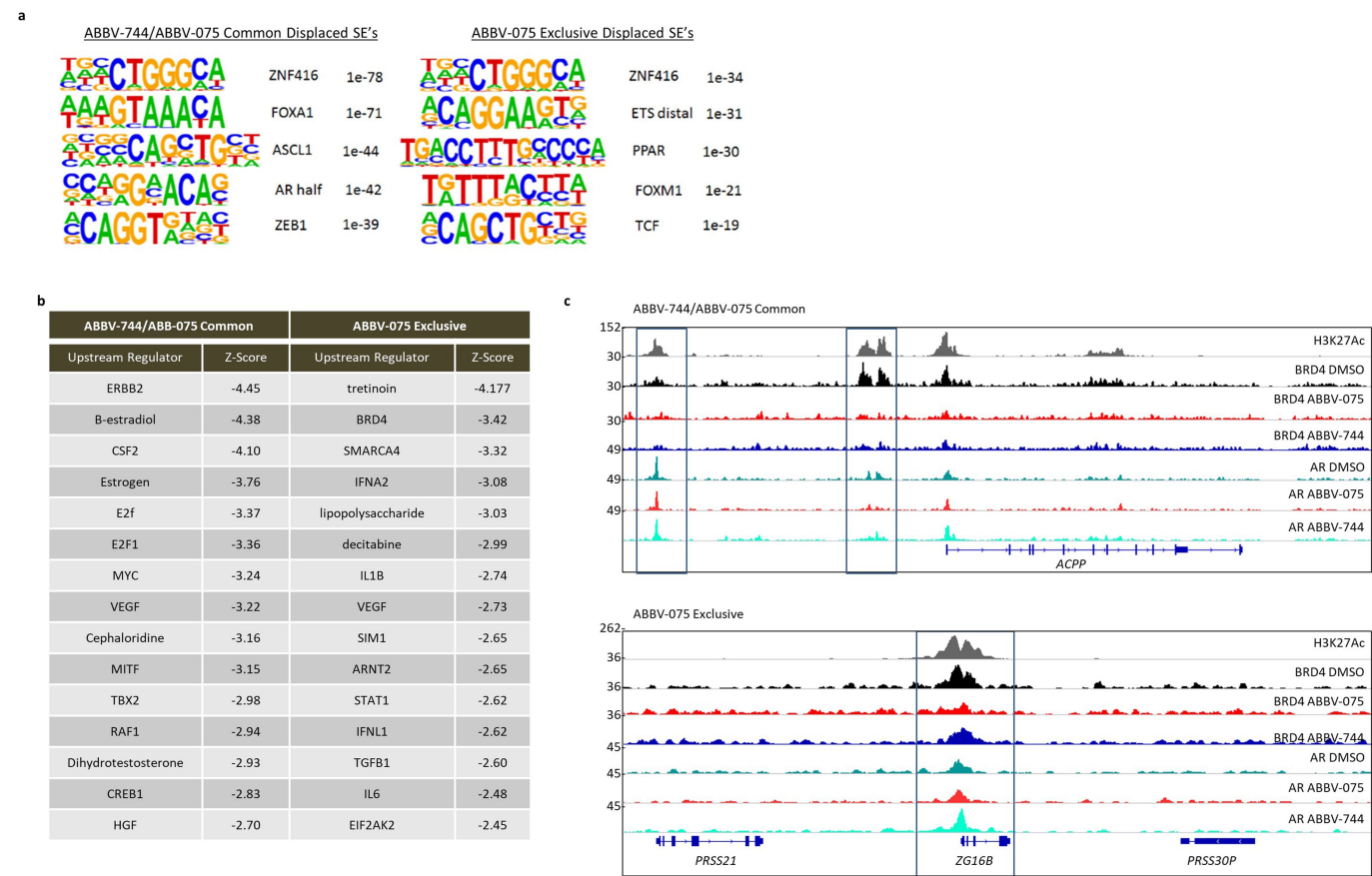
ABBV-075, ABBV-744 and BD2 and DbBis as described in the literature. **e**, Antiproliferative activities of ABBV-075 and ABBV-744 against IEC-6 and LNCaP cells and the activities of both compounds in a Mk-CFU assay—an assay that measures the generation of megakaryocytes from mouse haematopoietic stem cells—carried out by Stemcell Technology.



Extended Data Fig. 3 | ABBV-744 mimics enzalutamide and ABBV-075 to block AR-dependent transcription. **a**, Comparison of differentially regulated genes from this study with those reported in the literature using JQ1 and IBET. **b**, Reduction in MYC and KLK2 protein levels detected by western blot after treatment for 24 h with ABBV-075 (60 nM) or ABBV-744 (90 nM); no effect on AR was found. ABBV-075 but not ABBV-744 increases HEXIM1 protein levels. Representative of $n = 3$ independent experiments with similar results. For gel source data, see Supplementary Fig. 2. **c**, Biochemical, biophysical and cellular characteristics of the BD1 inhibitor (BD1i) described in the indicated GSK patent application. Bottom, Expression of *KLK2* and *MYC* in LNCaP cells after

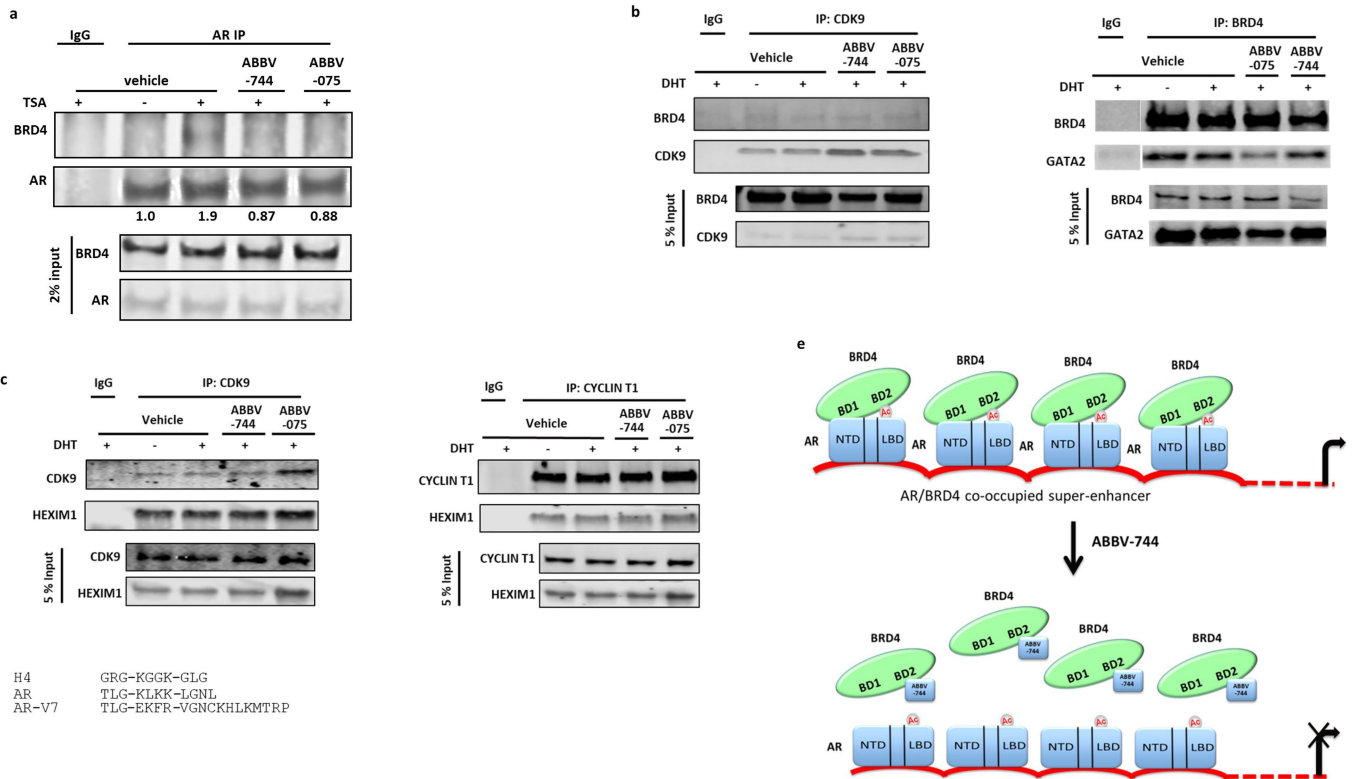
6 h treatment with ABBV-075 (60 nM), ABBV-744 (90 nM), BD1i (200 nM) or ABBV-744 (90 nM) and BD1i (200 nM) was determined by qPCR. Data are mean \pm s.d. ($n = 3$ biologically independent samples) and are representative of $n = 2$ independent experiments. **d**, Gene set enrichment analysis of RNA-seq data ($n = 2$) from LNCaP cells treated with ABBV-075, ABBV-744 or enzalutamide. Statistical significance was determined using a false-discovery rate (FDR) (Benjamini–Hochberg correction) and negative enrichment scores (NES) with $q < 0.05$ are listed in the table. Venn diagram shows the overlap of enriched hallmarks with each treatment. AR, MYC and E2F gene set enrichment analyses are shown as examples.





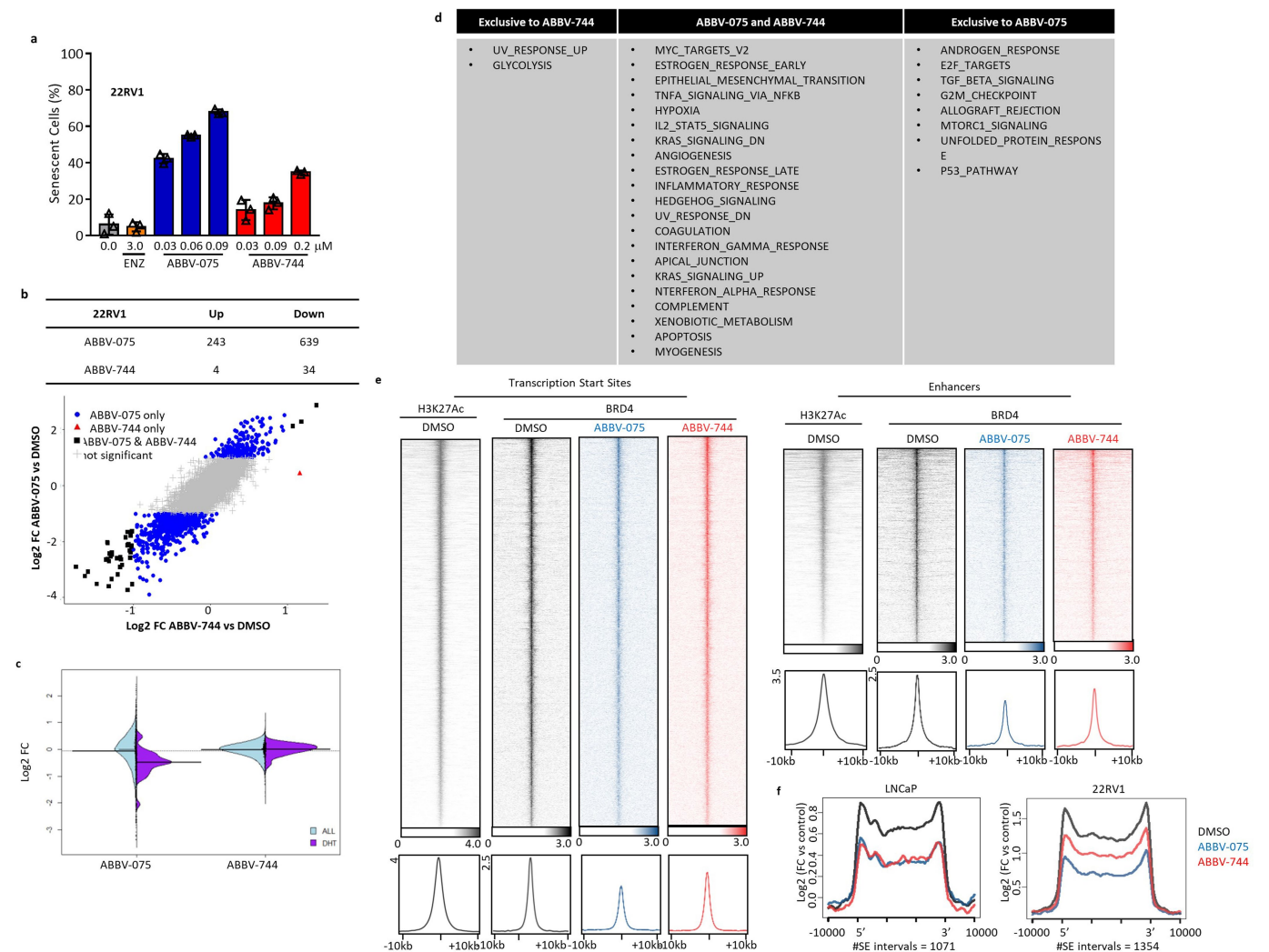
Extended Data Fig. 5 | BD2-dependent BRD4 binding motifs and upstream regulators. a, HOMER motifs enriched in super-enhancers in which ABBV-744 and ABBV-075 (common) displaced BRD4 or super-enhancers in which only ABBV-075 displaced BRD4 (exclusive), $n = 1$. Statistics were derived using FDR (Benjamini–Hochberg correction) and q values are shown. **b**, Upstream regulators for differentially expressed genes ($n = 2$) associated with ABBV-744

and ABBV-075 BRD4-displaced super-enhancers compared with ABBV-075-exclusive super-enhancers ($n = 1$), as analysed by ingenuity pathway analysis. AR, E2F1 and MYC all associated with common BRD4-displaced super-enhancers. **c**, Gene track examples of differential displacement pattern for ABBV-744 and ABBV-075 commonly sensitive (*ACPP*) or ABBV-075 exclusive (*ZG16B*).



Extended Data Fig. 6 | BD2-dependent BRD4-AR interaction. **a**, LNCaP cells were treated for 16 h with DHT in the presence of vehicle, ABBV-744 (90 nM) or ABBV-075 (60 nM) with or without trichostatin A (TSA) ($0.5 \mu\text{g ml}^{-1}$). AR immunoprecipitation (IP) using nuclear extracts pulled down BRD4 in trichostatin-A- and DHT-treated samples. ABBV-744 and ABBV-075 blocked BRD4 co-immunoprecipitation with AR. Fold change values from densitometry analysis are listed below the BRD4 blot, in which a 1.9-fold increase in the AR:BRD4 immunocomplex was measured in the trichostatin-A- and vehicle-treated lane compared with 0.87 or 0.88 after treatment with ABBV-744 or ABBV-075, respectively. Western blot of 2% immunoprecipitation input revealed no change in nuclear protein levels after inhibitor treatment. **b**, LNCaP cells were treated for 16 h with DHT in the presence of vehicle, ABBV-744

(90 nM) or ABBV-075 (60 nM). CDK9 or BRD4 immunoprecipitation using nuclear extracts pulled down BRD4 or GATA2, which is not blocked by treatment with ABBV-744. **c**, LNCaP cells were treated for 16 h with DHT in the presence of vehicle, ABBV-744 (90 nM) or ABBV-075 (60 nM). CDK9 or cyclin T1 immunoprecipitation using nuclear extracts pulled down HEXIM1, which is not blocked or enhanced by treatment with ABBV-744. **d**, Alignment of a KXXX motif in H4, AR and the lack of this motif in AR-V7. **e**, Cooperative interaction of BD1 and BD2 of BRD4 with acetylated AR at BRD4-AR co-occupied super-enhancers may underlie sensitivity to ABBV-744. **a-c**, Results are representative of $n > 2$ independent experiments. For **a-c** gel source data, see Supplementary Fig. 2.

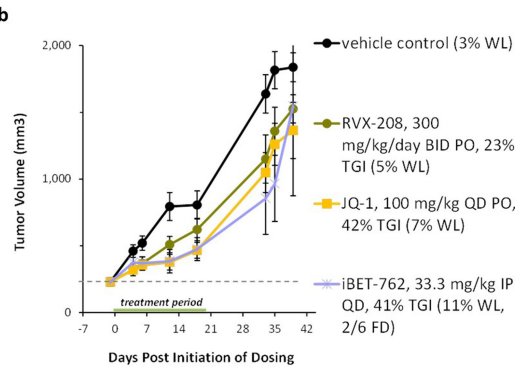


Extended Data Fig. 7 | 22RV1 cells are resistant to ABBV-744. **a**, ABBV-075 but not ABBV-744 induces a robust dose-dependent increase of senescent (β -galactosidase-positive) 22RV1 cells after 7 days of treatment. Data are mean \pm s.d. ($n = 3$ biological replicates) and are representative of $n = 2$ independent experiments. **b**, Scatter plot of gene expression changes ($n = 2$) caused by ABBV-075 (60 nM) or ABBV-744 (90 nM) treatment for 24 h in DHT-stimulated 22RV1 cells. Statistical analysis of fold change (FC) > 2.0 , $P < 0.01$ was conducted using the DESeq2 method. **c**, Split violin representation of DHT-regulated compared with all differentially expressed genes in 22RV1 from RNA-seq as shown in **b**. The long solid line represents the mean fold change. The small lines represent individual data points. The dotted line represents the

overall average. Statistical significance between all versus DHT was determined by two-tailed unpaired Student's t -test and $P < 0.01$ by DESeq2. ABBV-075 affects both DHT and a broad distribution of genes, whereas ABBV-744 has a more limited effect on both DHT-stimulated genes and overall. **d**, ABBV-075 but not ABBV-744 negatively regulated the androgen response in 22RV1 cells as shown by gene set enrichment analysis. NES > 2.0 , $q < 0.05$ calculated using FDR (Benjamini-Hochberg correction). **e**, H3K27Ac and BRD4 ChIP-seq heat maps at transcription start sites and enhancers in 22RV1 cells. **f**, ABBV-744 less effectively displaces BRD4 from super-enhancers in the resistant 22RV1 cell line compared with sensitive LNCaP cells.

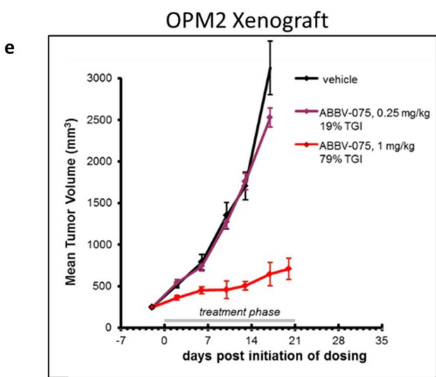
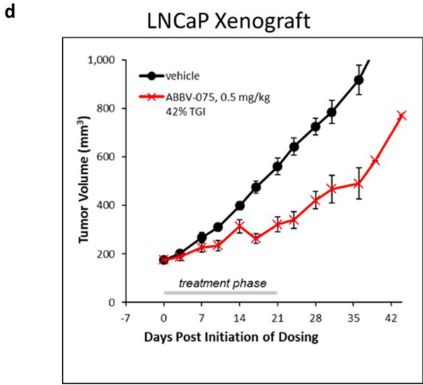
a

	Rat Tox Study				Mouse Efficacy Study		
	Dose (mg/kg)	AUC (μg.hr/ml)	Platelet reduction	Exposure multiple over efficacious exposure in mouse	Dose (mg/kg)	AUC (μg.hr/ml)	TGI
ABBV-075	3	3.66	59%	3 X	1	1.2	64%
ABBV-744	30	27.5	20%	25 X	4.7	1.1	64%



c

BET Inhibitor Efficacy comparison in LNCaP Model			
	TGI	Dose (mg/kg)	Comments
ABBV-744	64%	4.7	1/16 MTD
ABBV-075	64%	1	MTD
JQ1	42%	100	MTD
iBET	41%	33	MTD
RVX208	28%	300	Highest feasible dose, MTD not reached
ABBV-744	79%	75	MTD



Extended Data Fig. 8 | In vivo efficacy and tolerability of BD2 selective inhibitors and DbBis. a, Sprague-Dawley rats ($n = 3$ animals per group) were treated daily with vehicle, ABBV-075 (3 mg kg^{-1}) or ABBV-744 (30 mg kg^{-1}) for 14 days, and platelet counts were determined using the standard method. Efficacious exposure levels of ABBV-075 (1 mg kg^{-1}) and ABBV-744 (4.7 mg kg^{-1}) in mice and exposure levels associated with the indicated doses of each compound in rats were determined in separate pharmacokinetic studies using different animals ($n = 3$ animals per group). **b**, Antitumour activity of well-known BET inhibitors in the xenograft model in which LNCaP cells were

implanted in the mouse flank. JQ-1 and iBET-762 were administered at their respective MTD. RVX-208 was administered at its maximal achievable dose. Data are mean \pm s.e.m. of tumour size for each treatment group ($n = 6$). WL, maximum weight loss relative to initial value; FD, found dead. **c**, Efficacy comparison of BET inhibitors in the LNCaP model. **d**, **e**, Mice bearing LNCaP tumours (**d**; $n = 9$ per group) or OPM2 tumours (**e**; $n = 10$ per group) were treated with vehicle or ABBV-075 using oral gavage at the indicated amounts for 21 days (PO, QDX21). Data are mean \pm s.e.m. of tumour size for each treatment group.

Extended Data Table 1 | Data collection and refinement statistics

Data collection and refinement statistics of ABBV-744 in complex with BRD2/BD1 and BRD2/BD2

	BRD2-D2 ABBV-744	BRD2-D1 ABBV-744
Data collection		
Space group	P4 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	107.8, 107.8, 89.8	48.7, 56.1, 107.0
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	2.43 (2.43 - 2.48) *	1.97 (1.97 - 2.01) *
<i>R</i> _{pim}	0.046 (0.41)	0.059 (0.95)
<i>I</i> / σ <i>I</i>	14 (2.2)	16 (2.5)
Completeness (%)	100 (100)	97.6 (97)
Redundancy	6.6 (6.4)	6.4 (6.9)
Refinement		
Resolution (Å)	2.44	1.98
No. reflections	38380	20710
<i>R</i> _{work} / <i>R</i> _{free} (%)	21.1 / 23.8	19.8 / 23.3
No. atoms		
Protein	5436	1875
Ligand	216	72
Water	298	132
<i>B</i> -factors		
Protein	63	50
Ligand	58	44
Water	52	59
R.m.s. deviations		
Bond lengths (Å)	0.009	0.010
Bond angles (°)	0.93	0.88

Crystal structure coordinates and X-ray diffraction data of ABBV-744 in complex with BD1 of BRD2 and BD2 of BRD2 have been deposited in the Protein Data Bank with accession numbers 6E6J and 6ONY.
*Values in parentheses are for the highest-resolution shell.

Extended Data Table 2 | Antiproliferative activities of ABBV-744 across cancer cell lines

a

	Anti-proliferative IC ₅₀ (nM)						
Cell Line	LNCaP	MDA-PCa-2b	MDV-R	22RV1	VCaP	PC3	DU-145
AR Status	T878A	T878A, L702H	T878A, F876L	AR-V7	AR-V7	AR Negative	AR Negative
ABBV-744	11	9	18	467	354	>1000	>1000
Enzalutamide	550	>30,000	>30,000	>30,000	>30,000	>30,000	>30,000

b

Tumor Types	Cell Line	ABBV-075 IC ₅₀ (nM)	ABBV-744 IC ₅₀ (nM)		Tumor Types	Cell Line	ABBV-075 IC ₅₀ (nM)	ABBV-744 IC ₅₀ (nM)
AML	SIG-M5	2.8	2.1		OV	ES-2	6.0	830.0
AML	OCI-AML2	2.7	3.0		OV	Cov413B	17.0	880.0
AML	MV-4-11	2.6	3.1		TNBC	BT549	27.9	916.0
AML	EOL1	2.8	8.4		OV	OV56	21.0	916.0
AML	Kasumi1	3.0	13.0		NSCLC	NCI-H2347	45.0	968.0
AML	OCI-AML3	2.9	13.0		Head & Neck	FADU	22.0	>1000
AML	Nomo1	3.8	18.0		Neuroblastoma	Kelly	45.2	>1000
AML	HNT-34	4.4	58.0		Neuroblastoma	SKNF1	75.3	>1000
PC	LNCaP	3.8	10.8		NSCLC	NCI-H1792	41.0	>1000
PC	MDV-R	10.0	18.0		NSCLC	NCI-H727	87.0	>1000
PC	MDA-PCa-2b	13.5	38.3		NSCLC	NCI-H2170	100.0	>1000
TNBC	HCC2157	2.1	46.7		NSCLC	HCC-1395	108.0	>1000
TNBC	DU4475	9.0	110.3		NSCLC	NCI-H1563	117.0	>1000
TNBC	HCC1187	7.1	120.8		NSCLC	NCI-H827	988.0	>1000
TNBC	HS578T	5.4	139.7		NSCLC	NCI-H661	1000.0	>1000
OV	OVCAR3	8.0	167.0		NSCLC	NCI-H2935	1000.0	>1000
TNBC	MDA-MB-453	10.5	245.0		OV	OVCAR8	12.0	>1000
OV	A2780	16.0	290.0		OV	COV434	13.0	>1000
OV	PA-1	14.0	320.0		OV	CaoV-3	32.0	>1000
OV	OC314	10.0	348.0		PC	PC3	125.5	>1000
PC	VCaP	6.6	354.0		TNBC	HCC1806	14.6	>1000
TNBC	MDA-MB-468	31.0	454.1		TNBC	HCC38	25.9	>1000
PC	22RV1	16.5	467.0		TNBC	SUM149PT	61.0	>1000
TNBC	HCC1599	8.9	479.7		TNBC	CAL120	77.3	>1000
TNBC	MDA-MB-468	26.5	518.2		TNBC	HCC70	90.6	>1000
OV	SKOV3	17.0	600.0		TNBC	MDA-MB-231	138.2	>1000
OV	OVCAR5	14.0	696.0		TNBC	MDA-MB-436	160.1	>1000
PC	DU-145	130.2	706.9		TNBC	MDA-MB-157	292.7	>1000
NSCLC	NCI-H1703	48.0	806.0		TNBC	HCC1937	508.2	>1000
					TNBC	HCC1143	579.6	>1000

a, Segregation of ABBV-744 sensitivity with AR status in prostate cancer cell lines. **b**, ABBV-075 and ABBV-744 IC₅₀ values in a 5-day proliferation assay.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Illumina Genome Analyzer for sequence data, Commercial softwares (Studylog Systems, Inc., South San Francisco, CA) was used to collect in vivo tumor model data. Prestima software was used for in life and hematology data collection. Biacore T200 instrument and manufacturer provided software was used to collect SPR binding data. Envision plate reader with manufacturer supplied software was used to collect TR-FRET and NanoBRET data. Enspire plate reader with manufacturer supplied software was used to collect cell proliferation data.

Data analysis

Commercial software was used to analyze all data in this study as described in each section of the methods. These include Microsoft Excel, Prism GraphPad 5, Ingenuity Pathway Analysis, ArrayStudio, Biacore T200 software from manufacturer.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNASeq and ChIPSeq dataset can be accessed in GEO (Accession GSE118152, GSE118247, GSE130269). Crystal coordinates and X-ray diffraction data was deposited in the protein databank with the accession code 6E6J and 6ONY. Other datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For efficacy studies, a one sided t-test was used to determine the number of animals needed to obtain 80% power at alpha = 0.05. For rat tox studies, sample size of n=3 animals per group was based on internal experience of ability to identify test article related changes during drug candidate selection.
Data exclusions	No data were excluded from the analysis.
Replication	Experiments were repeated with same conditions and obtained similar results. The number of repeats were indicated in figure legends.
Randomization	For efficacy study, mice were randomized into treatment groups using Studylog software (Studylog Systems, Inc., South San Francisco, CA) based on tumor volume. For rat tox study, animal allocation to vehicle and treatment groups was at random based on body weight.
Blinding	Partial blinding for efficacy studies was used. A multiple technicians formulated and dosed compounds and randomized the groups. Additional investigators blinded to the test agents measured tumor volumes during the study. Toxicologic data analysis is generally performed in unblinded fashion which was the case for data described in this paper.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Information on all of the antibodies used in the study is presented in SI Table
Validation	H3K27Ac Ab noted on Active motif website to be modENCODE validated, NGS-QC certified, and validated for ChIP-Seq. BRD4 Ab is cited in at least 11 literature publications for ChIP and ChIP-Seq. AR Ab is cited in at least 28 literature publications including ChIP and ChIP-Seq applications. Antibody information is presented in SI Table.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The source and authentication of all eukaryotic cells in the study is presented in SI Table.
Authentication	Cell lines were authenticated using GenePrint 10 STR Authentication Kit (Promega, Madison, WI)
Mycoplasma contamination	Cell lines were tested for mycoplasma using MycoAlert Detection Kit (Lonza, Walkersville, MD) and all lines tested negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines used in this study

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	NSG-male mice (Jackson Laboratory), Fox Chase SCID [®] (Charles River Labs) mice, and Sprague Dawley (CrI:CD(SD)) rat strain from commercial sources were used. Male rat 56-58 days of age at initiation of testing article administration were used. NSG and Fox Chase SCID [®] male mice 6-8 weeks of age at time of study initiation were used.
Wild animals	No wild animals used in the study
Field-collected samples	No field-collected samples used in the study

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	Accession GSE118152, GSE118247
Files in database submission	<i>Provide a list of all files available in the database submission.</i>
Genome browser session (e.g. UCSC)	<i>Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.</i>

Methodology

Replicates	Each ChIP-Seq experiment was n=1.
Sequencing depth	All experiments were single end, 75 nt reads. For individual experiments total/usable: BRD4 DHT 39,825,247/22,844,767; BRD4DHT ABBV-744 37,353,171/18,060,657; BRD4 DHT ABBV-075 38,458,410/22,734,829; BRD4 DHT ENZ 34,719,339/21,325,188; AR 39,178,357/27,668,040; H3K27Ac 33,779,643/26,271,979.
Antibodies	Active Motif H3K27Ac cat#39133 lot 8, Bethyl BRD4 cat#A301-985A lot 6, Santa Cruz AR cat#sc-13062 lotB2616.
Peak calling parameters	Peaks were called using MACS2.1.0 narrow, pvalue cutoff 1e-7.
Data quality	Peaks that were on the ENCODE blacklist of known false ChIP-Seq peaks were removed.
Software	Illumina Casava 1.8 software used for basecalling. Reads were aligned to hg19 using BWA algorithm, USeq platform for Intersecting Regions and Neighboring Gene identifications (http://useq.sourceforge.net/). Further analysis of aligned bam files was done using NGSPlot (https://github.com/shenlab-sinai/ngsplot) to visualize heatmaps and generate average profile plots. NGSPlot provided heatmap and average profile plot figures.

Zucchini consensus motifs determine the mechanism of pre-piRNA production

<https://doi.org/10.1038/s41586-020-1966-9>

Received: 6 July 2019

Accepted: 25 November 2019

Published online: 29 January 2020

Natsuko Izumi^{1,5}, Keisuke Shoji^{1,2,5}, Yutaka Suzuki³, Susumu Katsuma⁴ & Yukihide Tomari^{1,3*}

PIWI-interacting RNAs (piRNAs) of between approximately 24 and 31 nucleotides in length guide PIWI proteins to silence transposons in animal gonads, thereby ensuring fertility¹. In the biogenesis of piRNAs, PIWI proteins are first loaded with 5'-monophosphorylated RNA fragments called pre-pre-piRNAs, which then undergo endonucleolytic cleavage to produce pre-piRNAs^{1,2}. Subsequently, the 3'-ends of pre-piRNAs are trimmed by the exonuclease Trimmer (PNLDC1 in mouse)^{3–6} and 2'-*O*-methylated by the methyltransferase Hen1 (HENMT1 in mouse)^{7–9}, generating mature piRNAs. It is assumed that the endonuclease Zucchini (MitoPLD in mouse) is a major enzyme catalysing the cleavage of pre-pre-piRNAs into pre-piRNAs^{10–13}. However, direct evidence for this model is lacking, and how pre-piRNAs are generated remains unclear. Here, to analyse pre-piRNA production, we established a Trimmer-knockout silkworm cell line and derived a cell-free system that faithfully recapitulates Zucchini-mediated cleavage of PIWI-loaded pre-pre-piRNAs. We found that pre-piRNAs are generated by parallel Zucchini-dependent and -independent mechanisms. Cleavage by Zucchini occurs at previously unrecognized consensus motifs on pre-pre-piRNAs, requires the RNA helicase Armitage, and is accompanied by 2'-*O*-methylation of pre-piRNAs. By contrast, slicing of pre-pre-piRNAs with weak Zucchini motifs is achieved by downstream complementary piRNAs, producing pre-piRNAs without 2'-*O*-methylation. Regardless of the endonucleolytic mechanism, pre-piRNAs are matured by Trimmer and Hen1. Our findings highlight multiplexed processing of piRNA precursors that supports robust and flexible piRNA biogenesis.

piRNAs are a class of small RNAs, approximately 24–31 nucleotides (nt) in size, produced from transposons and from discrete genomic loci called piRNA clusters¹⁴, and guide PIWI proteins to target transcripts. PIWI proteins possess an endonucleolytic activity, referred to as 'slicer', which directly cleaves target RNAs in the cytoplasm^{14–17}. In addition, a subset of PIWI proteins mediates transcriptional silencing in the nucleus^{18–20}. In germ cells, piRNA biogenesis is coupled with reciprocal slicing between complementary transcripts derived from transposons and piRNA clusters, a process called the ping-pong cycle^{14,15} (Extended Data Fig. 1). To generate mature piRNAs, PIWI proteins are first loaded with long single-stranded RNA fragments bearing a 5' monophosphate, called pre-pre-piRNAs^{1,2}. The pre-pre-piRNA is then endonucleolytically cleaved at a position 3' downstream of the PIWI-bound region to generate two cleavage fragments^{1,2,12,13}. In mice, silkworms and many other animals, the 5'-cleavage fragment, called a pre-piRNA, is shortened to the mature length by Trimmer (PNLDC1 in mouse), a PARN-like 3'-to-5' exonuclease localized on the mitochondrial surface^{3–6}, and 2'-*O*-methylated by the methyltransferase Hen1 (HENMT1 in mouse)^{7–9}. The 3' cleavage fragment is loaded into the next PIWI protein as a new pre-pre-piRNA. As a result, a series of 'trailing' pre-piRNAs are consecutively generated^{1,2,12,13} and matured by Trimmer and Hen1 (Extended Data Fig. 1).

The endonuclease Zucchini (MitoPLD or PLD6 in mouse)^{10,11}, which is localized on the mitochondrial outer membrane, is assumed to mediate cleavage of the PIWI-bound pre-pre-piRNAs^{12,13}. Because trailing piRNAs often start with a 5' uridine (U)^{2,4,12,13}, it is believed that cleavage activity of Zucchini has a preference for a site immediately before U *in vivo*. However, purified Zucchini protein shows nonspecific endoribonuclease activity *in vitro*^{10,11,21}. Owing to this discrepancy, the identity of the endonuclease for pre-pre-piRNAs remains unclear and is ambiguously and cautiously described in the literature^{1,2,22}. Thus, an *in vitro* system that can faithfully recapitulate the endonucleolytic reaction mediated by Zucchini is needed to resolve these ambiguities and discrepancies.

Two parallel pathways produce pre-piRNAs

To accurately investigate how pre-piRNAs are generated from pre-pre-piRNAs, it is necessary to block further processing of pre-piRNAs by Trimmer. In previous studies, knockdown of Trimmer in the silkworm cell line BmN4 resulted in only a slight extension of piRNA lengths^{3,21}, suggestive of residual Trimmer activity. To overcome this, we used CRISPR–Cas9 to generate Trimmer-knockout (Tri-KO) BmN4 cells (Extended Data Fig. 2a–c). We identified Tri-KO lines deficient in both

¹Laboratory of RNA Function, Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan. ²Department of Agrobiological and Bioresources, School of Agriculture, Utsunomiya University, Utsunomiya, Japan. ³Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan. ⁴Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. ⁵These authors contributed equally: Natsuko Izumi, Keisuke Shoji. *e-mail: tomari@iam.u-tokyo.ac.jp

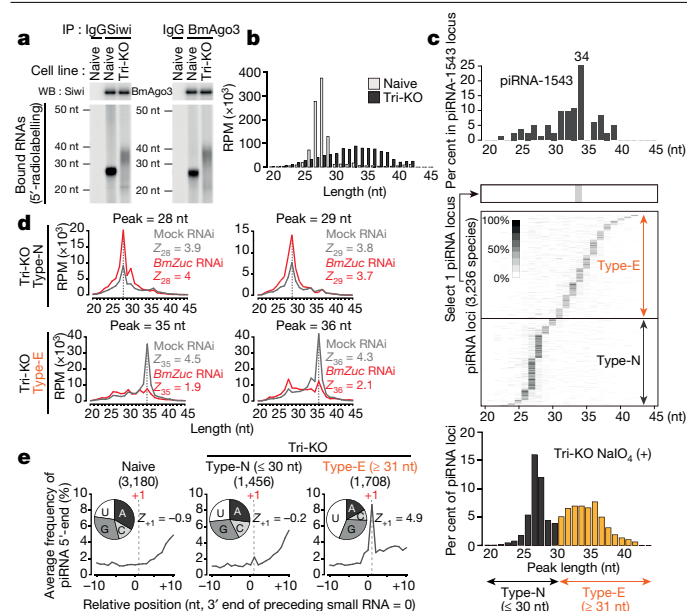


Fig. 1 | Two types of small RNAs accumulate in Tri-KO cells.

a, Immunoprecipitated (IP) Siwi or Bmago3 from naive or Tri-KO BmN4 cells was analysed by western blotting (WB) (top) and bound RNAs were detected by 5' radiolabelling (bottom). IgG, immunoprecipitation with non-immunized rabbit IgG. **b**, Length distribution of small RNAs mapped to 3,236 piRNA loci in the total small RNA library from naive or Tri-KO BmN4 cells. See also Extended Data Fig. 2i, j. **c**, The most abundant small RNA length among the reads sharing the same 5' end was defined as the peak length for each piRNA locus (for example, peak length = 34 nt for piRNA-1543, top). The 3,236 piRNA loci were aligned in the order of their peak lengths in the Tri-KO-NaIO₄ library (middle). piRNA loci with peak length of ≤ 30 nt were defined as type-N and those with peak length of ≥ 31 nt were defined as type-E (bottom). See also Extended Data Fig. 2k. **d**, Changes in the length distribution of NaIO₄-treated Tri-KO small RNAs bearing peak lengths of 28 or 29 nt (type-N), or 35 or 36 nt (type-E) caused by depletion of BmZuc. Mock indicates knockdown for *Renilla* luciferase. See also Extended Data Fig. 2l. Z_n denotes the z score at position n . RPM, reads per million. **e**, Mean occurrence of piRNA 5' ends relative to the peak position of each piRNA locus. Pie charts show the nucleotide composition immediately after the peak position of each piRNA locus. The numbers of analysed piRNA loci are shown in the parentheses. The per cent nucleotide composition in the silkworm genome corresponding to positions 11–45 of piRNA loci is 26:23:22:29 (T:G:C:A).

Trimmer protein and the in vitro trimming activity (Extended Data Fig. 2d–f). Tri-KO cells lacked mature 27–28 nt piRNAs and accumulated longer RNAs of about 30–40 nt (Extended Data Fig. 2g, red line) that co-immunoprecipitated with Siwi or Bmago3 (Fig. 1a). Overexpression of wild-type (WT) but not catalytically inactive Trimmer E30A (EA) recovered mature-length piRNAs (Extended Data Fig. 2h). These results suggest that silkworm pre-piRNAs are about 30–40 nt in length and are trimmed by Trimmer for maturation, irrespective of which PIWI protein they bind.

To characterize the pre-piRNAs in Tri-KO cells, we sequenced 20–50 nt small RNAs from Tri-KO cells with or without NaIO₄ treatment, which enables specific detection of 2'-O-methylated species. Small RNAs mapping to well-defined 3,236 piRNA loci³ showed a sharp distribution at 27–28 nt in naive BmN4 cells, but had a broad length distribution around 30–40 nt in Tri-KO cells (Fig. 1b). The small RNAs in Tri-KO cells were largely protected from NaIO₄ treatment (Fig. 1b, Extended Data Fig. 2i), suggesting that they are 2'-O-methylated, with longer species more efficiently methylated than shorter ones (Extended Data Fig. 2j). We determined the most frequent small RNA length (peak length) for each piRNA locus in the NaIO₄-treated library and plotted the peak lengths for all 3,236 piRNA loci without considering the small RNA

abundance from each locus (Fig. 1c, Extended Data Fig. 2k). In contrast to mature piRNAs in naive cells (Extended Data Fig. 2k), the peak lengths of Tri-KO small RNAs showed a clear bimodal distribution: one peak at 27–28 nt and a broader peak around 35 nt (Fig. 1c, bottom). For simplicity, we refer to the piRNA loci with peak length up to 30 nt as type-N (non-extended) and those with peak length greater than 30 nt as type-E (extended).

To test the requirement for *Bombyx mori* Zucchini (BmZuc) during processing of pre-pre-piRNAs into pre-piRNAs in silkworms, we knocked down BmZuc in Tri-KO cells. Notably, depletion of BmZuc did not affect the length distribution of type-N small RNAs, but strongly decreased peaks of length greater than 30 nt in the type-E small RNAs (Fig. 1d, Extended Data Fig. 2l and Supplementary Note 1), suggesting that BmZuc is required to produce pre-piRNAs from type-E loci. Supporting this idea, the genomic nucleotide immediately following the 3' end of type-E small RNAs in Tri-KO cells tended to be U (Fig. 1e), a proposed hallmark of Zucchini-mediated cleavage called the '+1U bias'^{2,4–6,12,13}. Moreover, type-E small RNAs in Tri-KO cells were frequently accompanied by immediately downstream piRNAs on the same genomic strand (Fig. 1e), a pattern typically observed in trailing piRNAs or pre-piRNAs^{2,4,12,13}. By contrast, these signatures were nearly absent in Tri-KO type-N small RNAs (Fig. 1e). Taken together, we conclude that BmZuc mediates the production of type-E pre-piRNAs, whereas type-N pre-piRNAs are generated via a BmZuc-independent pathway.

We next investigated whether BmZuc generates pre-piRNAs for both Siwi and Bmago3, the two PIWI proteins in silkworms. To this end, we first defined Siwi- and Bmago3-dominant piRNA loci (Extended Data Fig. 3a, see Methods). We then plotted the peak length of Tri-KO small RNAs separately for Siwi- or Bmago3-dominant piRNA loci, and observed similar bimodal distributions corresponding to type-N and type-E (Extended Data Fig. 3b). BmZuc depletion reduced the peak-length populations of type-E small RNAs for both Siwi- and Bmago3-dominant piRNA loci (Extended Data Fig. 3c), suggesting that BmZuc mediates type-E pre-piRNA production regardless of which PIWI protein is bound to the pre-pre-piRNA. However, compared with Siwi-dominant type-E pre-piRNAs, Bmago3-dominant type-E pre-piRNAs showed a weaker +1U bias (Extended Data Fig. 3d) and had a lower frequency of immediately downstream piRNAs (Extended Data Fig. 3e). Thus, even though BmZuc mediates pre-pre-piRNA cleavage for both Siwi and Bmago3, the production of downstream trailing piRNAs is largely restricted to Siwi.

We next investigated how pre-piRNAs in the type-N group are generated. In flies, most Ago3-bound piRNAs are processed from pre-piRNAs generated by piRNA-guided slicing at a downstream position²³ (Supplementary Discussion). To determine whether the 3' end of silkworm pre-piRNAs can be generated by downstream slicing, we analysed sense and antisense piRNAs mapped to the downstream region of type-N or type-E piRNA loci. The abundance of sense piRNAs in the downstream region was similar for type-N and type-E piRNA loci (Fig. 2a, sense strand). By contrast, antisense piRNAs at approximately 41–52 nt from their 5' ends were observed more frequently in the downstream region of type-N loci (Fig. 2a, antisense strand), for both Siwi-dominant and Bmago3-dominant loci (Extended Data Fig. 3f). Antisense piRNAs in this region can, in theory, guide slicing of pre-pre-piRNAs and generate the 3' end of 31–42 nt pre-piRNAs. Therefore, unlike in flies^{12,13,23}, downstream slicing of pre-pre-piRNAs in silkworms is probably determined by the context of the cleavage site and not by the identity of PIWI proteins. However, the peak lengths of Tri-KO small RNAs in the type-N group (less than 30 nt) were shorter than the expected pre-piRNA lengths based on the downstream slicing sites (31–42 nt) (Fig. 1c), implying that they are somehow fragmented into shorter species in Tri-KO cells.

To further investigate pre-piRNA generation, we examined four piRNA loci representing a Siwi-dominant (piRNA-1528) or Bmago3-dominant (piRNA-66) type-E locus, as well as a Siwi-dominant (piRNA-2986) or Bmago3-dominant (piRNA-304) type-N locus (Extended Data

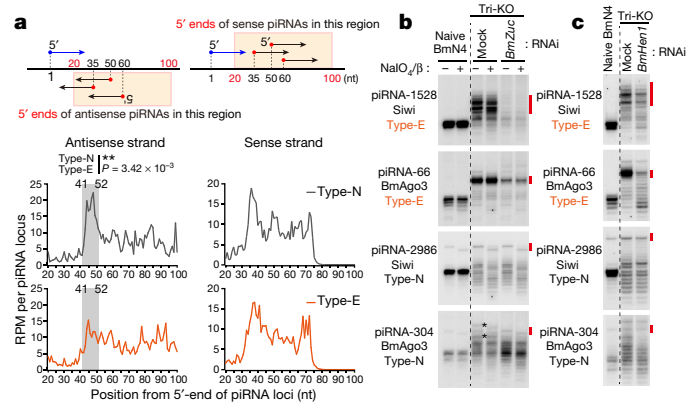


Fig. 2 | The 3' ends of type-E, but not type-N, pre-piRNAs are efficiently 2'-O-methylated. **a**, The 5' ends of piRNAs mapped to 20–100 nt downstream of piRNA loci were mapped on the antisense (left) or sense (right) genomic strand. Type-N piRNAs have more antisense piRNAs at 41–52 nt from the 5' ends than type-E piRNAs (two-sided Wilcoxon signed-rank test, $n = 12$). See also Extended Data Fig. 3f. **b**, **c**, Northern blot analysis of the four representative piRNAs in naive or Tri-KO BmN4 cells depleted of the indicated protein by RNAi. In **b**, total RNAs were treated with or without NaIO₄ elimination. Asterisks, BmZuc-dependent fragments. Red bars, putative pre-piRNAs. See also Extended Data Fig. 3g.

Fig. 3g). Type-N piRNA-2986 and 304, but not type-E piRNA-1528 and 66, have downstream ping-pong sites with readily detectable complementary piRNAs, which can guide PIWI-catalysed slicing of pre-pre-piRNAs. We examined small RNAs deriving from these loci in Tri-KO cells by northern blotting, and confirmed the accumulation of corresponding pre-piRNAs (Fig. 2b, mock, red lines). BmZuc depletion decreased type-E, but not type-N, pre-piRNA levels (Fig. 2b), reinforcing the idea that type-E pre-piRNAs are generated via BmZuc-mediated cleavage. Notably, type-N pre-piRNAs showed many shorter heterogeneous RNA fragments (Fig. 2b). These data suggest that pre-piRNAs generated by PIWI-catalysed slicing are intrinsically unstable and prone to non-specific degradation, at least in the absence of Trimmer. This could explain why the peak lengths of type-N small RNAs were 30 nt or shorter in Tri-KO cells (Fig. 1c). By contrast, type-E pre-piRNAs were more stable, especially pre-piRNA-66.

Given our observation that the 2'-O-methylation level was generally higher for longer small RNAs than shorter ones in Tri-KO cells (Fig. 1b and Extended Data Fig. 2i, j), we predicted that type-E pre-piRNAs are efficiently 2'-O-methylated. Indeed, type-E pre-piRNA-1528 and pre-piRNA-66 were refractory to NaIO₄ treatment (Fig. 2b). By contrast, type-N pre-piRNA-2986 and pre-piRNA-304, as well as their degradation products, were mostly—if not completely—shortened by one nucleotide by a NaIO₄ β -elimination reaction (Fig. 2b). Thus, type-E pre-piRNAs produced via BmZuc-mediated cleavage are more efficiently 2'-O-methylated than type-N pre-piRNAs generated by downstream piRNA-guided slicing. Consistently, type-E pre-piRNAs 1528 and 66 became prone to nonspecific degradation upon depletion of the 2'-O-methyltransferase BmHen1 in Tri-KO cells, whereas type-N pre-piRNAs 2986 and 304 and their degradation products, which are intrinsically poorly 2'-O-methylated, were largely unaffected (Fig. 2c). Mature piRNAs were fully 2'-O-methylated in naive BmN4 cells, regardless of how their pre-piRNAs are generated (Fig. 2b, Naive BmN4), supporting the model that 3'-end trimming by Trimmer is tightly coupled with 2'-O-methylation by BmHen1²⁴.

Type-N and type-E piRNA loci are heterogeneously distributed even within a single transposon (Extended Data Fig. 3h), suggesting that how pre-piRNAs are produced is determined at the level of individual piRNA loci. We also note that the separation between the type-N and type-E groups is not absolute; there are many cases in which pre-piRNAs

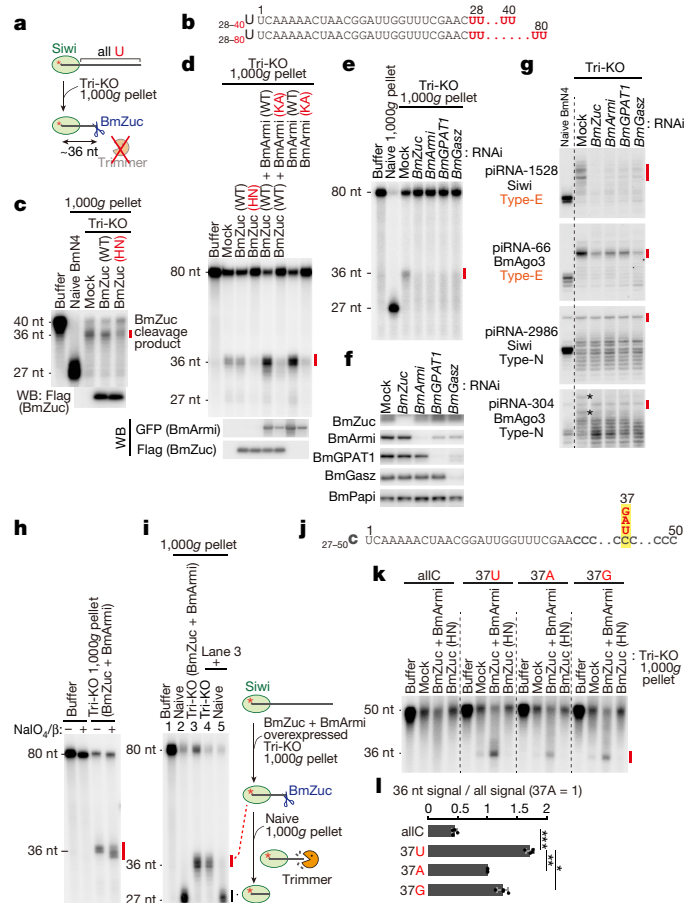


Fig. 3 | BmZuc requires BmArmi, BmGPAT1 and BmGasz for cleavage of Siwi-loaded pre-pre-piRNAs in vitro. **a**, Schematic of in vitro BmZuc cleavage assay. See also Extended Data Fig. 4a. **b**, RNA substrates used in **c–e**, **h** and **i**. **c–e**, Siwi-loaded 28–40 U RNA (top, **c**) or 28–80 U RNA (bottom; **d**, **e**) was incubated with 1,000g pellet from naive or Tri-KO cells overexpressing BmZuc(WT) or the catalytic mutant BmZuc(HN) (**c**, **d**) and/or BmArmi(WT) or the ATP-binding mutant BmArmi(KA) (**d**), or 1,000g pellet from Tri-KO cells depleted of the indicated protein by RNAi (**e**). The expression of Flag-tagged BmZuc and GFP-tagged BmArmi was confirmed by western blot (**c**, **d**, bottom). See also Extended Data Fig. 4c. **f**, Western blot analysis of the 1,000g pellet used in **e**. See also Extended Data Fig. 4e. **g**, Northern blot analysis of representative type-E or type-N pre-piRNAs in Tri-KO cells depleted of the indicated proteins by RNAi. Asterisks, BmZuc-dependent fragments. Red bars, putative pre-piRNAs. **h**, **i**, Siwi-loaded 28–80 U RNA was incubated with 1,000g pellet from Tri-KO cells overexpressing BmZuc and BmArmi. After incubation, RNAs were extracted and treated with NaIO₄ followed by β -elimination (**h**), or naive 1,000g pellet was added and further incubated (**i**). **j**, RNA substrates used in **k**. **k**, Siwi-loaded 27–50 C RNAs bearing U, A or G at position 37 were incubated with 1,000g pellet from Tri-KO cells overexpressing BmZuc and BmArmi or BmZuc(HN). **l**, Quantification of the 36-nt cleavage fragments produced by 1,000g pellet from Tri-KO cells overexpressing BmZuc and BmArmi in **k**. Data are mean \pm s.d. from four technically independent experiments. Bonferroni-corrected P values from two-sided paired t -tests are as follows: $^{*}P = 0.0163$; $^{**}P = 0.00106$; $^{***}P = 0.000485$.

are produced by both mechanisms, as represented by pre-piRNA-304 (Fig. 2b, asterisks) and pre-piRNA-1249 (Extended Data Fig. 3i).

In vitro analysis of BmZuc activity

We next sought to recapitulate pre-piRNA production in vitro. We previously established a cell-free system to monitor the 3'-end-trimming reaction by Trimmer using mitochondria-containing 1,000g pellets²⁴ (Extended Data Fig. 4a, left). Both Trimmer and BmZuc are

mitochondrial outer-membrane proteins, so we anticipated that the same strategy could be applied to detect BmZuc activity in the 1,000g pellet from Tri-KO cell homogenate (Fig. 3a, Extended Data Fig. 4a, right). Since Zucchini is thought to cleave 5' of U, we first used a series of single-stranded (ss)RNAs bearing a poly(U) sequence as model substrates (Fig. 3b). Incubation of Siwi-loaded 40-nt poly(U)-containing RNA with Tri-KO 1,000g pellet produced an RNA fragment of about 36 nt, which is much longer than the mature trimming product observed with naive 1,000g pellet. (Fig. 3c). The 36-nt fragment was also observed when we used an 80-nt poly(U)-containing RNA (Extended Data Fig. 4b, ATP+). For both 40- and 80-nt RNAs, overexpression of catalytically inactive BmZuc H141N (HN) decreased the ~36-nt signal (Fig. 3c, Extended Data Fig. 4b), suggesting that active BmZuc is required to generate this fragment. Thus, BmZuc catalyses the production of the 36-nt RNA fragment in vitro, regardless of the initial length of Siwi-loaded poly(U)-containing RNAs. This is consistent with the idea that the PIWI proteins themselves position Zucchini on pre-pre-piRNAs².

Depletion of ATP from the in vitro reaction abolished the 36-nt cleavage product (Extended Data Fig. 4b, ATP–), suggesting that BmZuc-mediated cleavage requires ATP. Purified Zucchini cleaves ssRNAs in an ATP-independent manner^{10,11,21}, whereas Armitage (MOV10L1 in mice)—a factor required for the biogenesis of trailing piRNAs—is an ATP-dependent RNA helicase^{25,26}. To examine whether BmArmi is required for BmZuc-mediated cleavage in vitro, we overexpressed wild-type BmArmi or its ATP binding mutant, K692A (KA), with or without BmZuc in Tri-KO cells. Overexpression of wild-type BmArmi alone strongly promoted in vitro cleavage of the Siwi-loaded 80-nt RNA, suggesting that BmArmi is a rate-limiting factor for BmZuc-mediated cleavage (Fig. 3d). By contrast, overexpression of BmArmi(KA) inhibited the cleavage reaction, indicating the importance of the ATPase activity (Fig. 3d). Knockdown of BmZuc or BmArmi in Tri-KO cells abolished the production of the 36-nt RNA fragment in vitro and biogenesis of endogenous type-E, but not type-N, pre-piRNAs, confirming their requirement for the cleavage reaction (Fig. 3e–g, Extended Data Fig. 4c). In addition to BmArmi, BmZuc-mediated cleavage required two other proteins localized on the mitochondrial surface, BmGPAT1 and BmGas2 (Fig. 3e–g, Extended Data Fig. 4c–e, Supplementary Discussion), homologues of which have been genetically implicated in Zucchini-mediated piRNA production in flies and mice^{12,27–30}.

We found that the BmZuc in vitro cleavage product of about 36 nt was at least partly resistant to NaIO₄ treatment, suggesting that it is protected by 2'-O-methylation (Fig. 3h). Thus, our in vitro system properly recapitulates BmZuc-mediated cleavage of pre-pre-piRNAs and the production of 2'-O-methylated type-E pre-piRNAs. Finally, we examined whether Trimmer can trim the cleavage product generated by BmZuc to produce mature piRNAs, recapitulating processing in vivo. The 36-nt BmZuc cleavage product was efficiently converted into 27–28-nt mature piRNAs by naive 1,000g pellet, which contains endogenous Trimmer (Fig. 3i). This result validates the stepwise 3'-end maturation mechanism of type-E piRNAs: BmZuc cleaves pre-pre-piRNAs to generate pre-piRNAs, which are trimmed to the mature length by Trimmer.

Previous genetic and deep-sequencing analyses have suggested that Zucchini preferentially cleaves immediately 5' to U in vivo^{2,4–6,12,13}. We also observed that the 3' ends of type-E pre-piRNAs in Tri-KO cells have a modest +1U bias, especially for type-E pre-piRNAs bound to Siwi (Fig. 1e, Extended Data Fig. 3d). However, previous biochemical analyses using purified Zucchini proteins and naked RNAs have failed to detect this U preference^{10,11,21}. We applied our new in vitro system using mitochondria-containing pellets and Siwi-loaded pre-pre-piRNAs to revisit this inconsistency. We performed the in vitro BmZuc cleavage assay with a 50-nt RNA bearing a poly(C) sequence as well as variants that substituted the C at position 37 with U, A or G (Fig. 3j). Compared with the 37A, 37G and 37C RNAs, the 37U RNA substrate yielded moderately but significantly increased levels of the 36-nt cleavage product,

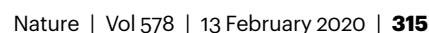
in a manner dependent on the catalytic activity of BmZuc (Fig. 3k, l). Thus, our system recapitulates the U preference of BmZuc, consistent with our bioinformatics analysis of type-E pre-piRNAs (Fig. 1e, Extended Data Fig. 3d).

BmZuc motif dictates piRNA biogenesis

The moderate U preference of BmZuc was apparent for poly(C)-based sequences in vitro (Fig. 3j–l). However, BmZuc does not always cleave immediately 5' to U in our in vitro system (Extended Data Fig. 4f). Moreover, natural type-E pre-piRNAs showed only a modest +1U bias (Fig. 1e, Extended Data Fig. 3d). Thus, the proposed U preference does not fully explain how the cleavage site is chosen by BmZuc. To investigate the substrate specificity of BmZuc in a comprehensive and unbiased manner, we performed a screen in Tri-KO cells. In brief, we constructed a plasmid-based library that expresses 35 nt of random sequence flanked by target sites for an abundant BmAgo3-dominant piRNA³¹ (Fig. 4a, see Methods and Supplementary Note 2). The transcripts are expected to be sliced by the BmAgo3-dominant piRNA, loaded into Siwi via the ping-pong pathway as new pre-pre-piRNAs, and cleaved by BmZuc within the downstream randomized region (Fig. 4a), producing various type-E pre-piRNAs. We first sequenced the library-derived small RNAs and examined their peak length distribution (Extended Data Fig. 5a). We observed library-derived small RNAs around 35 nt, recapitulating the size range of endogenous type-E pre-piRNAs. These approximately 35-nt RNAs were enhanced by overexpression of BmZuc and BmArmi, and strongly inhibited by BmZuc(HN), indicating that they are generated by BmZuc-mediated cleavage. We then aligned them at the 3' ends of their peak length (that is, putative BmZuc cleavage sites, defined as position 0) and analysed the nucleotide frequencies at each position. Focusing on the six nucleotides with the highest frequencies, we identified a sequence motif (–10A, –2A, –1U, 0U, +1U, +4C) (Fig. 4b), which was also consistently observed in the BmZuc + BmArmi overexpression condition.

To further investigate the 'BmZuc motif' in Siwi-bound pre-pre-piRNAs, we analysed two representative sequences from the library (84497 and 111750) that contain all six consensus nucleotides. We generated a series of mutants that alter the consensus sequence and performed the BmZuc cleavage assay (Fig. 4c, d, Extended Data Fig. 5b, c). Wild-type sequences showed site-specific cleavage at 34 or 35 nt, as expected from the in-cell screen. By contrast, 'All mut' sequences, in which all the six consensus nucleotides were mutated, lacked site-specific cleavage (Fig. 4d, Extended Data Fig. 5c), providing further supporting evidence that this motif determines the BmZuc cleavage site. Unexpectedly, mutating only the +1U, the proposed hallmark of Zucchini-mediated cleavage, did not inhibit BmZuc-mediated site-specific cleavage of these sequences (Fig. 4d, Extended Data Fig. 5c, +1U mut.). However, mutating –1U and 0U together strongly inhibited the cleavage at the correct position, whereas mutating –10A, –2A and +4C together had a minor effect. In sum, our findings reveal a previously unrecognized consensus motif that is important for BmZuc to precisely determine the cleavage site.

BmZuc generates the 3' ends for both Siwi- and BmAgo3-loaded pre-piRNAs (Extended Data Fig. 3b, c). To investigate whether BmZuc has a different nucleotide preference for pre-pre-piRNAs bound to BmAgo3, we constructed a reciprocal plasmid library whose transcripts were loaded into BmAgo3 as pre-pre-piRNAs with a randomized sequence. We then performed co-immunoprecipitation with Siwi from the Tri-KO cells transfected with the original library and co-immunoprecipitation with BmAgo3 from the cells transfected with the reciprocal library, and analysed the bound small RNAs with peak lengths of 31–44 nt after NaIO₄ treatment (Extended Data Fig. 5d, e). As expected, the small RNAs immunoprecipitating with Siwi showed very similar nucleotide preferences around their 3' ends as the BmZuc motif identified by the non-immunoprecipitation experiment using the same plasmid library



which endonucleolytic mechanism they are cleaved (Extended Data Fig. 5i and Supplementary Discussion). Thus, the sequence context of pre-pre-piRNAs has a major role in determining how silkworm pre-piRNAs are produced.

Discussion

Here we show that silkworm pre-piRNAs are generated by two parallel endonucleolytic mechanisms: BmZuc-mediated cleavage and PIWI-catalysed slicing (Fig. 4h). This multiplexed system supports robust and flexible piRNA biogenesis in silkworms, which have only two PIWI proteins (Supplementary Discussion). Regardless of how pre-piRNAs are generated and to which PIWI protein they are bound, Trimmer is essential for the maturation of piRNAs; this appears to be true also in mice and many other species² (Supplementary Discussion). It has been reported that Zucchini preferentially cleaves immediately 5' to U^{2,4-6,12,13}, highlighting the +1U bias as the one (and only) signature of Zucchini-mediated cleavage. Our data suggests that +1U alone is insufficient to determine the BmZuc cleavage sites and instead reveals previously unrecognized consensus motifs preferred by BmZuc (Fig. 4, Extended Data Fig. 6 and Supplementary Discussion). How these specific motifs are recognized warrants future investigation. Given that isolated Zucchini proteins do not show any apparent nucleotide specificity^{10,11,21}, we speculate that it is not Zucchini itself, but rather a 'reaction platform' on the mitochondrial surface formed by proteins such as Armitage and Gasz that have important roles in determining the cleavage site (Supplementary Discussion).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1966-9>.

- Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
- Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K. & Zamore, P. D. Single mechanism of biogenesis, initiated and directed by PIWI proteins, explains piRNA production in most animals. *Mol. Cell* **71**, 775–790 (2018).
- Izumi, N. et al. Identification and functional analysis of the pre-piRNA 3' Trimmer in silkworms. *Cell* **164**, 962–973 (2016).
- Ding, D. et al. PNLD1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. *Nat. Commun.* **8**, 819 (2017).
- Zhang, Y. et al. An essential role for PNLD1 in piRNA 3' end trimming and male fertility in mice. *Cell Res.* **27**, 1392–1396 (2017).

- Nishimura, T. et al. PNLD1, mouse pre-piRNA trimmer, is required for meiotic and post-meiotic male germ cell development. *EMBO Rep.* **19**, e44957 (2018).
- Horwich, M. D. et al. The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
- Kirino, Y. & Mourelatos, Z. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* **13**, 1397–1401 (2007).
- Saito, K. et al. Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* **21**, 1603–1608 (2007).
- Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**, 279–283 (2012).
- Nishimasu, H. et al. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**, 284–287 (2012).
- Han, B. W., Wang, W., Li, C., Weng, Z. & Zamore, P. D. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* **348**, 817–821 (2015).
- Mohn, F., Handler, D. & Brennecke, J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* **348**, 812–817 (2015).
- Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
- Gunawardane, L. S. et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
- De Fazio, S. et al. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**, 259–263 (2011).
- Reuter, M. et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**, 264–267 (2011).
- Aravin, A. A. et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31**, 785–799 (2008).
- Kuramochi-Miyagawa, S. et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917 (2008).
- Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).
- Nishida, K. M. et al. Hierarchical roles of mitochondrial Papi and Zucchini in *Bombix* germline piRNA biogenesis. *Nature* **555**, 260–264 (2018).
- Ge, D. T. et al. The RNA-binding ATPase, Armitage, couples piRNA amplification in Nuage to phased piRNA production on mitochondria. *Mol. Cell* **74**, 982–995 (2019).
- Hayashi, R. et al. Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature* **539**, 588–592 (2016).
- Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3' end formation of PIWI-interacting RNAs in vitro. *Mol. Cell* **43**, 1015–1022 (2011).
- Vourekas, A. et al. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev.* **29**, 617–629 (2015).
- Pandey, R. R. et al. Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in *Drosophila* ovaries. *PLoS Genet.* **13**, e1006956 (2017).
- Handler, D. et al. The genetic makeup of the *Drosophila* piRNA pathway. *Mol. Cell* **50**, 762–777 (2013).
- Muerdter, F. et al. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol. Cell* **50**, 736–748 (2013).
- Shiromoto, Y. et al. GPAT2, a mitochondrial outer membrane protein, in piRNA biogenesis in germline stem cells. *RNA* **19**, 803–810 (2013).
- Vagin, V. V. et al. Minotaur is critical for primary piRNA biogenesis. *RNA* **19**, 1064–1077 (2013).
- Shoji, K., Suzuki, Y., Sugano, S., Shimada, T. & Katsuma, S. Artificial "ping-pong" cascade of PIWI-interacting RNA in silkworm cells. *RNA* **23**, 86–97 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Cell culture, plasmid transfection, and generation of stable or knockout cell line in BmN4 cells

BmN4 cells (provided by T. Kusakabe, Kyushu University; not authenticated and not tested for mycoplasma contamination) were cultured at 27 °C in IPL-41 medium (AppliChem) supplemented with 10% fetal bovine serum. For plasmid transfection, 5–7.5 µg of plasmid DNAs were transfected into BmN4 cells (2.5×10^6 cells per 10 cm dish) with X-tremeGENE HP DNA Transfection Reagent (Sigma). For generation of stable cells expressing GFP-BmArmi, BmN4 cells were transfected with a GFP-tagged BmArmi expression vector and selected under 10 µg/ml puromycin for 3 weeks. For generation of Trimmer knockout cell line, BmN4 cells were co-transfected with pIEx1-MychCas9NLS expression vector and pBS-BmU6-sgTrimmer expression vector. One week later, the cells were reseeded at a low density ($\sim 1\text{--}4 \times 10^4$ cells per 15 cm dish) and cultured in 50–75% conditioned medium. About 3 weeks later, colonies were picked up under a microscope.

Plasmid construction

pIEx1-Trimmer WT and the catalytic mutant (E30A) were described previously³. The primer sequences for plasmid construction are listed in Supplementary Table 1.

pIEx1-MychCas9NLS. A DNA fragment coding Myc-hCas9-NLS was amplified from pRB14 (a gift from K. Förstemann)³² and cloned into pIEx-1 vector (Millipore/Novagen) by In-Fusion HD cloning kit (Takara Clontech).

pBS-BmU6-sgTrimmer. To generate pBS-BmU6-BbsI-chiRNA vector, the fly U6 promoter in pBS-U6-BbsI-chiRNA expression vector (a gift from K. Förstemann)³² was replaced with the *Bombyx mori* U6 promoter³³ amplified from the BmN4 genome. Synthesized DNA oligos for Trimmer sgRNA were annealed and inserted into BbsI-digested pBS-BmU6-BbsI-chiRNA vector.

pIExZ-BmZuc WT, H141N. To generate pIExZ vector, the ampicillin resistant gene and its promoter sequence in pIEx-1 vector (Millipore/Novagen) were replaced by *ie2* promoter and Zeocin resistant gene amplified from pIZ/V5-His vector (Thermo Fisher/Invitrogen). To enhance the expression, BmZuc coding sequence was codon-optimized to *Bombyx mori* using EMBOSS Backtranseq (http://www.ebi.ac.uk/Tools/st/emboss_backtranseq/). The BmZuc coding sequence was synthesized by GeneArt Strings DNA Fragments service (Life Technologies) and cloned into pIExZ vector. The catalytic mutant BmZuc (H141N) was generated by site-directed mutagenesis.

EGFP-BmArmi K692A. The ATP-binding mutant EGFP-BmArmi K692A was generated by site-directed mutagenesis into EGFP-BmArmi (a gift from T. Kusakabe and T. Tatsuke)³⁴.

Antibodies and western blotting

Rabbit anti-Siwi, anti-BmAgo3, anti-BmZuc, anti-BmArmi, anti-BmGPAT1, anti-BmGasz antibodies were generated by immunizing N-terminally His-tagged recombinant Siwi (aa 2–100), BmAgo3 (aa 2–100), BmZuc (aa 28–206), BmArmi (aa 2–294), BmGPAT1 (aa 602–870), BmGasz (aa 2–269) respectively (Scrum). The sera were affinity-purified by a column containing the immobilized recombinant protein. Anti-Trimmer and anti-BmPapi antibodies were described previously³. Anti-Flag (M2) (Sigma), anti-actin (Santa Cruz, sc-1616) and anti-GFP (B-2) (Santa Cruz) antibodies were purchased. Chemiluminescence was induced by Luminata Forte Western HRP Substrate (Millipore) and images were acquired by Amersham Imager 600 (GE Healthcare).

In vitro processing assay

In vitro ssRNA loading and trimming, NaO₄-mediated oxidation, and β-elimination were performed essentially as described previously²⁴.

Each substrate ssRNA was 5'-radiolabelled with T4 polynucleotide kinase (Takara) and [γ -³²P]ATP (PerkinElmer). For BmZuc cleavage assay, Tri-KO cells were resuspended in hypotonic buffer (10 mM HEPES-KOH (pH 7.4), 10 mM KCl, 1.5 mM MgCl₂, 1 mM DTT, 1× Complete EDTA-free protease inhibitor (Roche)) and incubated on ice for 20 min. Subsequently, the cell suspension was vortexed for 30 s, centrifuged at 1,000g for 20 min at 4 °C, and the supernatant was removed. The pellet was resuspended in hypotonic buffer and used as the 1,000g pellet fraction. Typically, 7 µl of the resuspended 1,000g pellet fraction was added to immunopurified Flag-Siwi-ssRNA complex on beads together with 3 µl of 40× reaction mix (containing ATP, ATP regeneration system, and RNase inhibitor)³⁵ and incubated at 25 °C for 2.5 h (Fig. 3c, e and Extended Data Fig. 4b, c) or 30 °C for 20 min (Fig. 3k), 2 h (Fig. 4d and Extended Data Fig. 5c), 2.5 h (Fig. 3h, i), or 3 h (Fig. 3d and Extended Data Fig. 4f). For BmZuc cleavage assay in an ATP-depleted condition (Extended Data Fig. 4b), hypotonic buffer was added instead of 40× reaction mix. For standard trimming assay, Flag-Siwi-ssRNA complex on beads was incubated with 1,000g pellet from naive BmN4 cells together with 40× reaction mix at 25 °C for 20 min (Extended Data Fig. 2f) or 30 °C for 1.5 h (Fig. 3i). In all the in vitro cleavage/trimming assays, lysates with an equal protein concentration were used in each experimental set. Images were acquired by Typhoon FLA 7000 (GE Healthcare) and analysed using Multi Gauge 3.0 (Fujifilm).

RNAi in BmN4 cells

For dsRNA preparation, template DNAs were prepared by PCR using primers containing T7 promoter listed in Supplementary Table 1. dsRNAs were transcribed using T7 Scribe Standard RNA IVT Kit (Cell Script) and purified with MEGAclear Transcription Clean-Up Kit (Thermo Fisher/Invitrogen). For dsRNA transfection, 5 µg of dsRNAs were transfected into BmN4 cells (6×10^5 cells per 10 cm dish) with X-tremeGENE HP DNA Transfection Reagent (Sigma). dsRNAs were repeatedly transfected every 3 days for four times.

Immunoprecipitation

For Siwi and BmAgo3 immunoprecipitation, cells were resuspended in buffer A (25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1.5 mM MgCl₂, 0.2% sodium deoxycholate, 0.1% lithium dodecyl sulfate, 0.4% NP-40, 0.5 mM DTT, 1× Complete EDTA-free protease inhibitor (Roche)) and incubated on ice for 20 min. The cell suspension was diluted with equal volume of buffer A without detergents and centrifuged at 17,000g for 30 min at 4 °C. The supernatant was incubated with normal rabbit IgG (Cell Signaling), anti-Siwi or anti-BmAgo3 antibody at 4 °C for 1 h, and then Dynabeads Protein G (Thermo Fisher/Invitrogen) was added. After incubation at 4 °C for 1 h, the beads were washed with buffer B (25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1.5 mM MgCl₂, 0.1% sodium deoxycholate, 0.05% lithium dodecyl sulfate, 0.2% NP-40, 0.5 mM DTT, 1× Complete EDTA-free protease inhibitor (Roche)). The immunoprecipitated proteins were eluted with SDS sample buffer, and bound RNAs were purified with mirVana miRNA Isolation Kit (Thermo Fisher/Invitrogen) for small RNA library preparation or TRI Reagent (Molecular Research Center) for 5' radiolabelling.

Genome extraction, RNA extraction, quantitative real-time PCR and northern blotting

Genomic DNA of BmN4 cells was extracted using NucleoSpin Tissue (Macherey-Nagel). Total RNAs prepared by TRI Reagent (Molecular Research Center) were used for real-time PCR and northern blotting. One microgram of total RNAs was reverse transcribed by PrimeScript RT reagent kit with gDNA eraser (Takara), and qRT-PCR was performed using KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) and the Thermal Cycler Dice Real Time System (Takara). For northern blotting, 10–12 µg of total RNAs were resolved by 15% urea polyacrylamide gel electrophoresis (PAGE) and transferred to Hybond-N membrane (GE Healthcare). After chemical crosslinking³⁶, 5' labelled antisense

Article

DNA probes were hybridized with Perfect Hyb Plus (Sigma) at 42 °C overnight. The primer sequences for genomic PCR, real-time PCR and DNA probes for Northern blotting are listed in Supplementary Table 1.

Immunofluorescence

Stable cells expressing GFP-tagged BmArmi were treated with 100 nM Mitotracker Red CMXRos (Cell Signaling) at 27 °C for 1 h. After fixing with 4% paraformaldehyde at room temperature for 10 min, the cells were permeabilized with 0.3% Triton X-100 for 5 min and incubated with PBS supplemented with 1% BSA (Sigma) and 0.1% Triton X-100 at room temperature for 1 h. Then, the cells were incubated with anti-BmGasZ antibody (1:400) in PBS supplemented with 1% BSA (Sigma) and 0.1% Triton X-100 at 4 °C overnight. Alexa Fluor 647 donkey anti-rabbit IgG antibody (Thermo Fisher/Invitrogen) was used as the secondary antibody. Images were captured using Olympus FV3000 confocal laser scanning system with a $\times 60$ oil immersion objective lens (PLAPON 60XO, NA 1.42, Olympus) and processed FV31S-SW Viewer software and Adobe Photoshop Elements 10.

Construction of plasmid-based randomized sequence library and sequencing

A DNA fragment containing an N35 random sequence embedded between two piRNA target sites was amplified by PCR using 3rand-const primers and synthetic N35-containing DNA oligos (IDT) (Siwi/BmAgo3-3endRAND50-double) as the template (Supplementary Table 1). The PCR products were digested with BamHI and HindIII, and cloned into the BamHI/HindIII sites of pEx-4 vector (Millipore/Novagen). The plasmid-based library expresses 35-nt of random sequence flanked by target sites for an abundant BmAgo3- (Fig. 4a, Extended Data Fig. 5d left, and Supplementary Note 2) or Siwi-dominant piRNA (Extended Data Fig. 5d right)³¹. The library was transfected into (1) Tri-KO cells (mock) or Tri-KO cells overexpressing either (2) wild-type BmZuc and BmArmi, to enhance BmZuc-mediated cleavage, or (3) the catalytic mutant BmZuc (HN), to repress BmZuc-mediated cleavage. Transcripts derived from the library are expected to be sliced by complementary piRNAs bound to BmAgo3/Siwi, loaded into Siwi/BmAgo3 via the ping-pong pathway, and cleaved by BmZuc within the downstream randomized region. We sequenced 20–50 nt small RNAs bearing the common sequence in their 5' region, and restored the original sequences downstream of the obtained small RNAs by using the sequence data of the plasmid library. The variation of the randomized region was estimated to be 215,879 for Siwi and 178,735 for BmAgo3 based on the number of distinct sequences with RPM > 0.25 in the reference libraries. The randomized sequence libraries for reference were constructed by PCR using a common reverse primer (plasmid-Hind-RandR) and a specific forward primer (plasmid-Siwi-RandF-index12 or plasmid-Ago3-RandF-index19, shown in Supplementary Table 1). The libraries were sequenced by the Illumina HiSeq 3000 platform to obtain 100-nt paired-end reads using a custom primer containing the consensus sequence (random plasmid sequence primer, shown in Supplementary Table 1) and an index read sequence primer provided by the manufacturer.

Small RNA library preparation

Small RNA libraries were prepared from 20–50 nt total, Siwi-bound, or BmAgo3-bound RNAs, according to the Zamore lab's open protocol (<https://www.dropbox.com/s/r5d7aj3hhyaborq/>)³⁷ with some modifications. The 3' adapter was conjugated with amino CA linker instead of dCC at the 3' end (GeneDesign) and adenylated using 5' DNA adenylation kit at the 5' end (NEB). To reduce a ligation bias, four random nucleotides were included in the 3' and 5' adapters [(5'-rAppNNNTGGAATTCTCGGGTGCCAAGG/amino CA linker-3') and (5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNN-3')] and the adapter ligation was performed in the presence of 20% PEG-8000³⁸, except for Fig. 1d and Extended Data Fig. 2l, 3c, 5a. After the 3' adapter ligation at 16 °C for ≥ 16 h, RNAs were size-selected by urea PAGE. In Fig. 1d, and

Extended Data Fig. 2l, 3c, 5a, the 3' and 5' adapters without the four random nucleotides were used. In Fig. 4e and Extended Data Fig. 5e, the 5' adapter without the four random nucleotides was used. For RNA extraction from polyacrylamide gel, ZR small-RNA PAGE Recovery Kit (ZYMO Research) was used. For small RNA library preparation from the randomized sequence library (Fig. 4e, Extended Data Fig. 5a, e), specific forward primer (piRNA-Siwi or BmAgo3-RandF, shown in Supplementary Table 1) was used in PCR to selectively amplify the plasmid-derived transcripts. Small RNA libraries were sequenced using the Illumina HiSeq 4000 platform to obtain 50-nt single-end reads.

Sequence analysis of endogenous small RNAs

After removal of adapter sequences by cutadapt³⁹, 20–45 nt reads without any ambiguous bases were mapped to sequences of defined piRNA loci³ with Bowtie⁴⁰ allowing one mismatch. Sam files were converted to bam files by SAMtools⁴¹ and then to bed files by BEDTools⁴². Length and 5'-end position for each piRNA were obtained from bed files using custom R programs. To determine Siwi- and BmAgo3-dominant loci, Siwi-immunoprecipitated and BmAgo3-immunoprecipitated libraries treated with NaIO₄ from naive BmN4 cells were compared and defined Siwi-dominant piRNA loci (RPM (Siwi-IP) > RPM (BmAgo3-IP), $n = 1,946$) and BmAgo3-dominant piRNA loci (RPM (BmAgo3-IP) > RPM (Siwi-IP), $n = 1,259$) (Extended Data Fig. 3a). For mouse small RNA analysis, piRNA loci were defined from the deep sequencing data by Gainetdinov et al.². In brief, 1st–23rd sequence of each read was extracted and the frequency of each 1st–23rd sequence was calculated in each library (SRR7760309, SRR7760310, SRR7760317, SRR7760318, SRR7760321, SRR7760322, SRR7760343, SRR7760344, SRR7760347, SRR7760348, SRR7760369, SRR7760370, SRR7760373, SRR7760374, SRR7760377, SRR7760378). Sequences that are abundantly found (RPM > 10) in at least one library, 36,431 in total, were selected and mapped to the mouse genome (GRCm38.p5) to define representative piRNA loci in pachytene spermatocytes.

Sequence analysis of randomized plasmids and small RNAs

Sequences in the randomized region of the plasmids were extracted and the frequency of each distinct sequence was calculated. Those sequences with RPM > 4 (80,966 species for Siwi and 70,442 species for BmAgo3) were used as references. For analysing small RNAs derived from the randomized libraries, adapters were trimmed by cutadapt³⁹ and 20–45 nt reads without any ambiguous bases but with a 15-nt common sequence from the plasmids were mapped to the randomized sequence references with Bowtie⁴⁰ allowing no mismatch. Sam files were converted to bam files by SAMtools⁴¹ and then to bed files by BEDTools⁴². Length and 5'-end position for each small RNA were obtained from bed files using custom R programs. To analyse the nucleotide frequencies relative to the 3' end of small RNAs, we selected sequences with peak lengths of 31–44 nt, corresponding to the size range of type-E pre-piRNAs.

Calculation of the similarity scores with weighted BmZuc motifs

The top 15 highest-frequency nucleotides in the randomized Siwi-immunoprecipitated or BmAgo3-immunoprecipitated libraries (Fig. 4e) were chosen to define the 'weighted BmZuc motif' of 17-nt long (from -12 to +4, 0 = predicted BmZuc cleavage site) for Siwi or BmAgo3 (Extended Data Fig. 5g, upper), with each nucleotide having a score of the log₂ value of the nucleotide frequency at that position, normalized to the randomized sequence references. The similarity score was calculated by summing up the weighted BmZuc motif if a nucleotide at a given position matches to the nucleotide at the corresponding position in the defined BmZuc motif for Siwi or BmAgo3 (Extended Data Fig. 5g, lower), by sliding the 17-nt window on each 27-nt sequence in the extracted genomic sequence pools or the control shuffled pool, using custom R programs. The control shuffled sequences have the average nucleotide composition of the silkworm genome

corresponding to positions 11–45 of piRNA loci (U[T]:G:C:A = 25.7:23.4:21.8:29.1)⁴³. For Fig. 4f, 27-nt genomic sequence pools corresponding to positions 19–45 nt of 1,946 Siwi-dominant and 1,259 BmAgo3-dominant piRNA loci (Extended Data Fig. 3a) were extracted and used to calculate the similarity score between the weighted BmZuc motif for Siwi or BmAgo3 and each extracted genomic sequence from the Siwi- or BmAgo3-dominant piRNA loci by a sliding window approach (Extended Data Fig. 5g, lower).

Statistics and reproducibility

Experiments in Figs. 1a, 3c–f, Extended Data Figs. 2d–g, 4d, f were independently performed twice with similar results. Experiments in Figs. 2b, c, 3g–i, 4d and Extended Data Figs. 2b, h, 3i, 4b, c, 5c were performed once. For the statistical analyses in Figs. 2a, 3l, Extended Data Figs. 3f, 5i, detailed statistical values were summarized in Supplementary Table 2. To estimate of effect sizes in Fig. 3l, the cohensD function in the lsr package was used. To estimate of effect sizes in Fig. 2a and Extended Data Figs. 3f, 5i the wilcoxsign_test and wilcox_test function in the coin package was used. No statistical methods were used to predetermine sample size. All experiments were not randomized and no blinding was used during data analysis.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The sequencing data reported in this paper are publicly available in DDBJ, under the accession number DRA008549. All other data are available from the authors upon reasonable request.

Code availability

All code required for bioinformatics analysis in this paper is available at https://github.com/kshoji-nt/BmZuc_cleavage.

32. Böttcher, R. et al. Efficient chromosomal gene modification with CRISPR/Cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic Acids Res.* **42**, e89 (2014).
33. Zhu, L., Mon, H., Xu, J., Lee, J. M. & Kusakabe, T. CRISPR–Cas9-mediated knockout of factors in non-homologous end joining pathway enhances gene targeting in silkworm cells. *Sci. Rep.* **5**, 18103 (2015).
34. Patil, A. A. et al. Characterization of Armitage and Yb containing granules and their relationship to nuage in ovary-derived cultured silkworm cell. *Biochem. Biophys. Res. Commun.* **490**, 134–140 (2017).
35. Haley, B., Tang, G. & Zamore, P. D. In vitro analysis of RNA interference in *Drosophila melanogaster*. *Methods* **30**, 330–336 (2003).

36. Pall, G. S. & Hamilton, A. J. Improved Northern blot method for enhanced detection of small RNA. *Nat. Protoc.* **3**, 1077–1084 (2008).
37. Fu, Y., Wu, P. H., Beane, T., Zamore, P. D. & Weng, Z. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* **19**, 531 (2018).
38. Kim, H. et al. Bias-minimized quantification of microRNA reveals widespread alternative processing and 3' end modification. *Nucleic Acids Res.* **47**, 2630–2640 (2019).
39. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*. **17**, 10–12 (2011).
40. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
41. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
43. Kawamoto, M. et al. High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* **107**, 53–62 (2019).
44. Kamminga, L. M. et al. Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J.* **29**, 3688–3700 (2010).
45. Simon, B. et al. Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure* **19**, 172–180 (2011).
46. Tian, Y., Simanshu, D. K., Ma, J. B. & Patel, D. J. Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc. Natl Acad. Sci. USA* **108**, 903–910 (2011).
47. Lim, S. L. et al. HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. *PLoS Genet.* **11**, e1005620 (2015).
48. Homolka, D. et al. PIWI slicing and RNA elements in precursors instruct directional primary piRNA biogenesis. *Cell Rep.* **12**, 418–428 (2015).
49. Yang, Z. et al. PIWI slicing and EXD1 drive biogenesis of nuclear piRNAs from cytosolic targets of the mouse piRNA pathway. *Mol. Cell* **61**, 138–152 (2016).

Acknowledgements We thank T. Kusakabe and T. Tatsuke for providing BmArmi expression vectors, K. Förstemann for providing hCas9 and sgRNA expression vectors, T. Kiuchi for sharing unpublished data and helpful discussion, and K. Kiyokawa and T. Horiuchi for technical assistance. A part of Illumina sequencing was performed in the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 ODO18174 Instrumentation Grant. We also thank Life Science Editors for editorial assistance, and P. Zamore and members of the Tomari laboratory for critical comments on the manuscript. This work was in part supported by a Grant-in-Aids for Scientific Research on Innovative Areas (grant 26113007 to Y.T.) from the Ministry of Education, Culture, Sports, Science and Technology in Japan and a Grant-in-Aid for Scientific Research (S) (grant 18H05271 to Y.T.), Grant-in-Aid for Scientific Research (B) (grant 16KT0064 to Y.S. and S.K.), Grant-in-Aid for Scientific Research on Innovative Areas (grant 17H06431 to S.K.), Grant-in-Aid for Young Scientists (B) (grant 17K17673 to N.I.), Grant-in-Aid for Scientific Research (C) (grant 19K06484 to N.I.), and a Grant-in-Aid for JSPS Fellows (grant 17J02408 to K.S.).

Author contributions N.I., K.S. and Y.T. conceived and designed the experiments and wrote the manuscript. N.I. performed biochemical experiments. K.S. performed bioinformatics analyses. Y.S. and S.K. supervised the bioinformatics analyses. Y.T. supervised the research. All the authors discussed the results and approved the manuscript.

Competing interests The authors declare no competing interests.

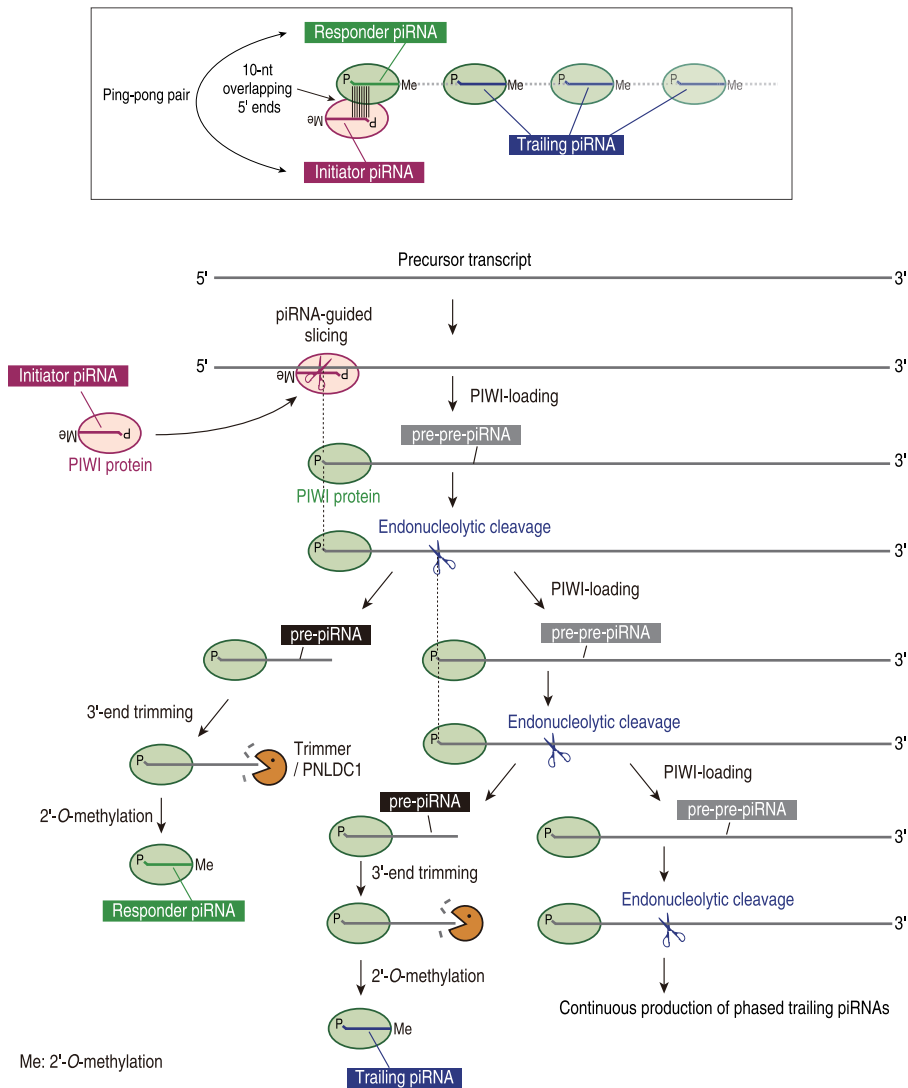
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1966-9>.

Correspondence and requests for materials should be addressed to Y.T.

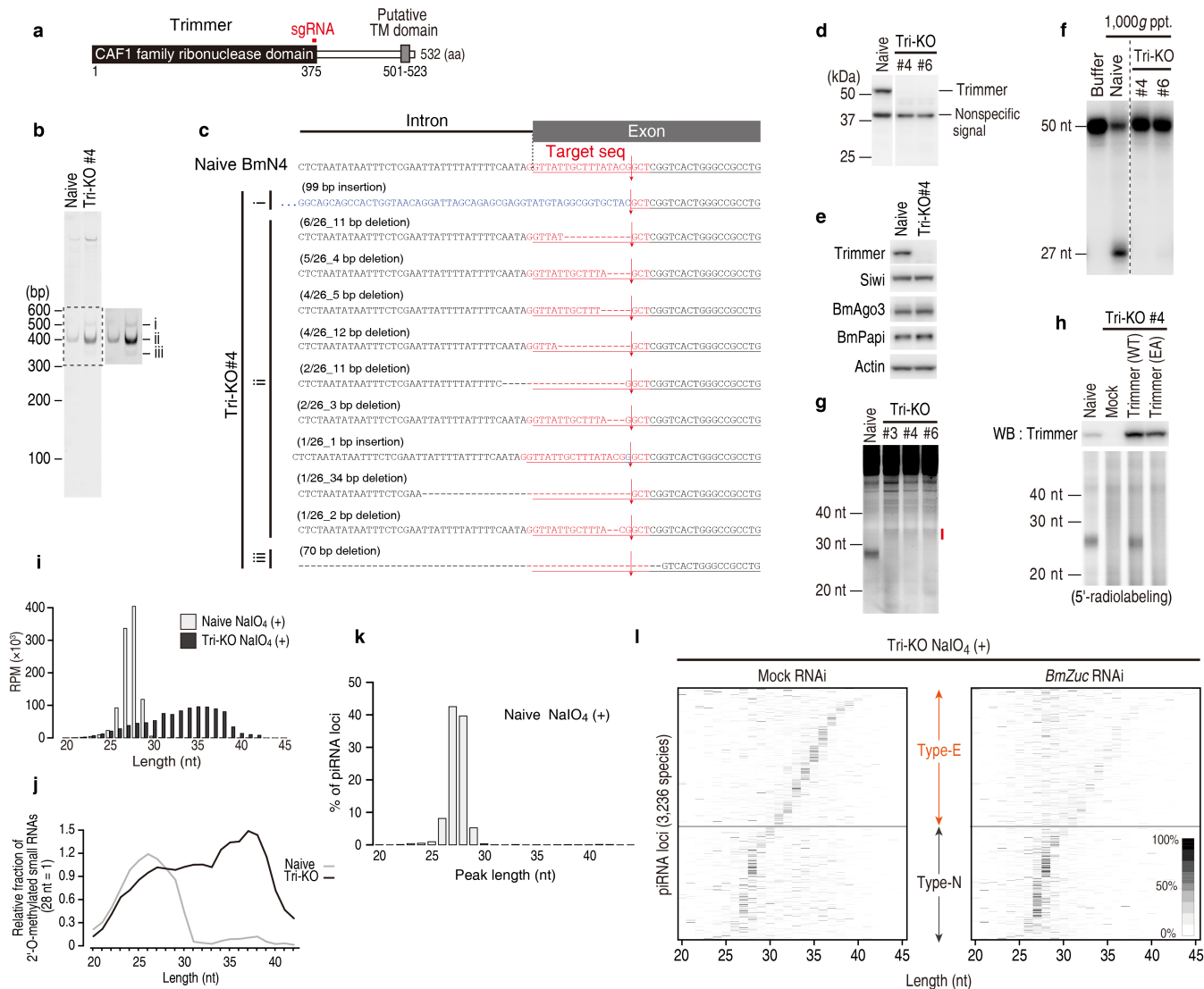
Peer review information Nature thanks René Ketting and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



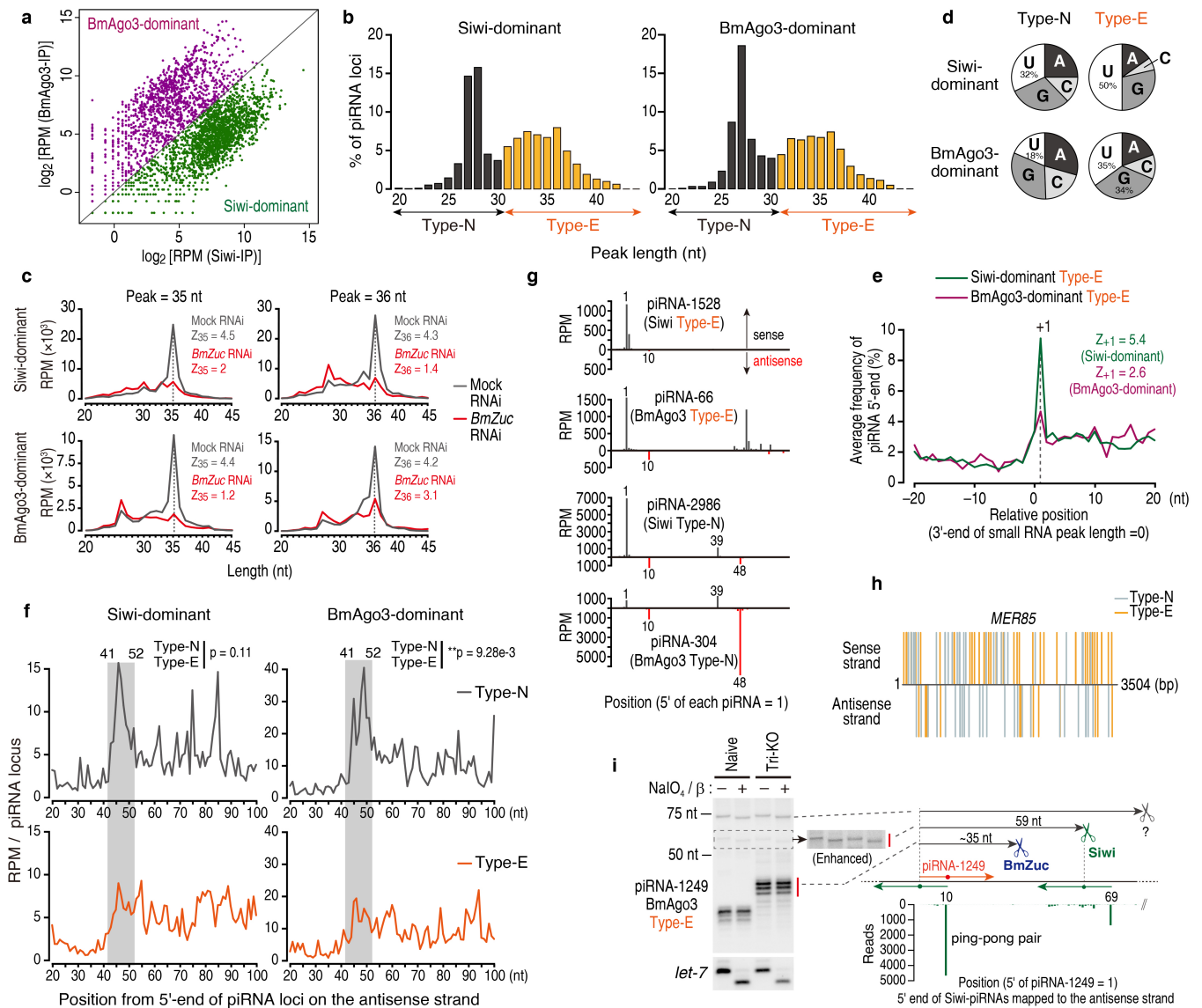
Extended Data Fig. 1 | The current model for the piRNA biogenesis initiated by the piRNA-guided PIWI-catalysed slicing in animal germ cells. The ping-pong cycle produces pairs of piRNAs ('initiator' and 'responder' piRNAs) that show 10-nt overlapping at their 5' ends as well as multiple 'trailing' piRNAs downstream of the responder piRNAs (top)¹. The ping-pong cycle is initiated by the slicing of a precursor transcript by an initiator piRNA-loaded PIWI protein. The PIWI-cleaved 5'-monophosphorylated fragment is handed over to a corresponding PIWI protein as a pre-pre-piRNA. Then, the PIWI-loaded pre-pre-piRNA is endonucleolytically cleaved at a downstream position^{1,2,12,13}. The resultant 5' cleavage fragment, called a pre-piRNA, is further trimmed by the 3'-to-5' exonuclease Trimmer (PNLDC1 in mouse)³⁻⁶ to the mature length, 2'-O-methylated by the methyltransferase Hen1 (HENMT1 in mouse)⁷⁻⁹, and becomes a responder piRNA (left pathway). Hen1-mediated 2'-O-methylation protects

mature piRNAs from degradation and tightens their binding to PIWI proteins^{7,44-47}. The 3' endonucleolytic cleavage fragment of the pre-pre-piRNA is loaded into a next PIWI protein as a new pre-pre-piRNA and endonucleolytically cleaved again at a downstream position, producing a new PIWI-loaded pre-piRNA^{1,2,12,13}. This pre-piRNA is then processed by Trimmer and Hen1 at the 3' end into a mature trailing piRNA (middle pathway). The 3' cleavage fragment of the second endonucleolytic cleavage is also loaded into a next PIWI protein, serving as a new pre-pre-piRNAs. As a result, a series of trailing piRNAs are consecutively produced downstream of the responder piRNA (right pathway)^{1,2,12,13,48,49}. These two piRNA biogenesis pathways lead to target-dependent amplification of piRNAs (via the ping-pong cycle) and expansion of piRNA sequences (via trailing piRNA production).



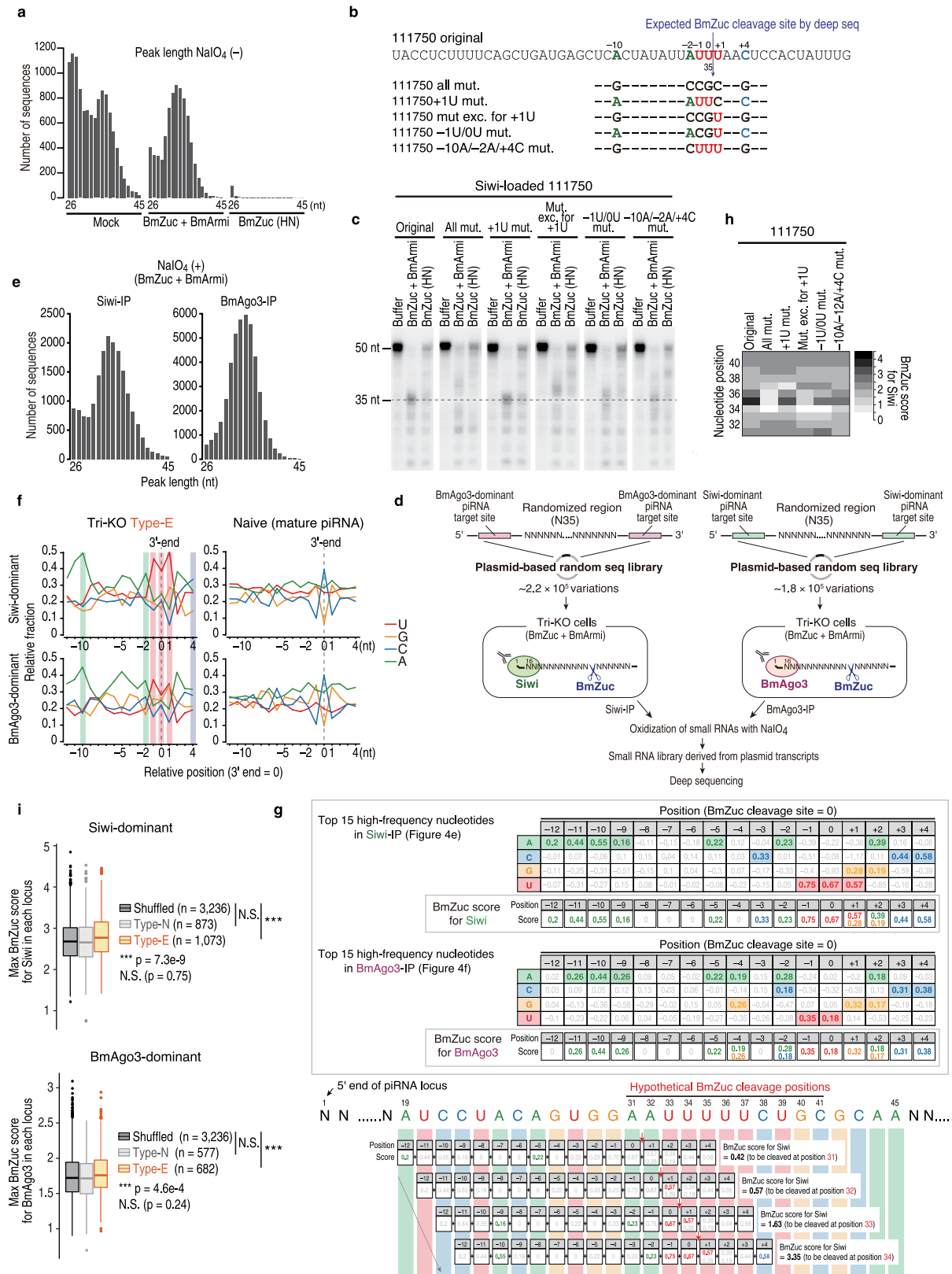
Extended Data Fig. 2 | Generation and characterization of Trimmer-knockout cells. **a**, Schematic representation of the domain structure of Trimmer and the position of the sgRNA target site for CRISPR-Cas9. **b**, Genomic PCR of a region including the sgRNA target site. In addition to the main PCR product (ii), two additional PCR products (i and iii) were detected only in Tri-KO#4 cells. Detailed genome sequences around the sgRNA target site in naive or Tri-KO#4 BmN4 cells. Genomic sequencing revealed various mutations at the sgRNA target site, suggesting a polyploid nature of the *trimmer* locus and/or imperfect cell cloning. **d**, Western blot analysis of Trimmer in two different Tri-KO cell lines (#4 and #6). Tri-KO line #4 was used in this study. **e**, Western blot analysis of whole-cell lysate from naive or Tri-KO#4 BmN4 cells. **f**, In vitro trimming assay for Siwi-loaded IU50 RNA using 1,000g ppt. from naive or two different Tri-KO cell lines. ppt., pellet.

g, SYBR Gold staining of total RNAs from naive or three different Tri-KO cell lines (#3, #4 and #6). **h**, Total RNAs extracted from Tri-KO #4 cells overexpressing wild-type Trimmer (WT) or its catalytic mutant E30A (EA) were 5' radiolabelled and detected by phosphor imaging. Mock indicates transfection of a control plasmid. Trimmer expression was analysed by western blotting (upper). **i**, Length distribution of small RNAs mapped to 3,236 piRNA loci in NaIO₄-treated small RNA library from naive or Tri-KO BmN4 cells. **j**, Relative fraction of 2'-O-methylated Tri-KO small RNAs in each length. **k**, Peak length distribution of piRNAs mapped to 3,236 piRNA loci in the NaIO₄-treated library from naive BmN4 cells. **l**, Changes by the depletion of BmZuc in the length distribution of Type-N (lower) or Type-E (upper) NaIO₄-treated small RNAs in Tri-KO cells.



Extended Data Fig. 3 | BmZuc is required to produce type-E pre-piRNAs for both Siwi and BmAgo3, whereas trailing piRNA production is largely restricted to Siwi. **a**, Scatter plot showing normalized piRNA abundance co-immunoprecipitated with Siwi or BmAgo3 from naive BmN4 cells for each piRNA loci. Green dots, Siwi-dominant piRNA loci ($n=1,946$); purple dots, BmAgo3-dominant piRNA loci ($n=1,259$). **b**, Peak length frequency of Tri-KO small RNAs for Siwi-dominant (left) or BmAgo3-dominant (right) piRNA loci. **c**, Length distribution of Tri-KO small RNAs bearing the peak length of 35 or 36 nt (type-E) for Siwi-dominant (upper) or BmAgo3-dominant (lower) piRNA loci. BmZuc knockdown abolished small RNAs with the peak lengths. Z_n denotes the z-score at position n (c, e). **d**, Siwi-dominant type-E pre-piRNAs show a stronger +1U preference than BmAgo3-dominant ones. **e**, Siwi-dominant type-E pre-piRNAs show a greater tendency to have downstream trailing piRNAs than BmAgo3-dominant ones. **f**, The 5' ends of piRNAs mapped to 20–100 nt downstream of type-N (top) or type-E (bottom) Tri-KO small RNAs

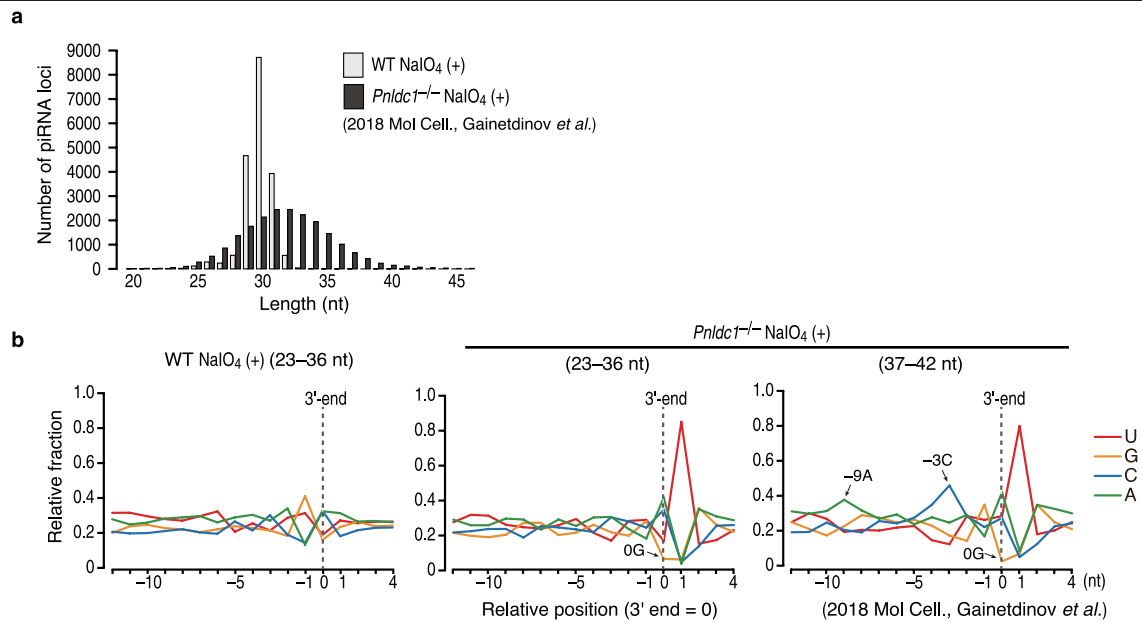
were mapped on the antisense strand, separately for Siwi-dominant (left) and BmAgo3-dominant (right) piRNA loci. Type-N piRNAs have more antisense piRNAs at -41–52 nt from the 5' ends than type-E piRNAs, regardless of which PIWI protein they bind (two-sided Wilcoxon signed rank test, $n=12$). **g**, Four representative type-E (piRNA-1528 and 66) and type-N (piRNA-2986 and 304) piRNA loci and their downstream genomic regions were mapped with the 5' ends of sense (grey) and antisense (red) piRNAs. **h**, Distribution of type-E and type-N piRNAs mapped to a transposon called MER85. **i**, An example of mixed modes of pre-piRNA production. Pre-pre-piRNA-1249 contains a BmZuc cleavage site and a slicing site by an antisense piRNA-loaded PIWI protein. The -35 nt BmZuc cleavage product, but not the 59 nt slicing product, is 2'-O-methylated. An unmethylated -75 nt fragment, which is possibly produced by another antisense piRNA-guided slicing, locates in an unannotated genomic region and cannot be assigned.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Calculation of BmZuc scores for Siwi or BmAgo3 based on the randomized sequence library analysis. **a**, Peak length distribution of small RNAs derived from the randomized sequence library. **b**, RNA substrates used in **c**. The top 6 nucleotides in the BmZuc motif are shown in colour and their mutations are shown in black. **c**, Siwi-loaded 111750-derived RNAs were incubated with Tri-KO 1,000g ppt. overexpressing BmZuc and BmArmi, or BmZuc(HN). Each gel image was adjusted to equalize the loading signal. ppt., pellet. **d**, Schematic representation of the randomized sequence library analysis for Siwi- or BmAgo3-loaded pre-piRNAs cleaved by BmZuc. **e**, Peak length distribution of Siwi- or BmAgo3-bound 2'-O-methylated small RNAs derived from the corresponding randomized sequence library. For Siwi immunoprecipitation, the same plasmid library as in **a** was used. **f**, Nucleotide composition around the 3' ends of mature piRNAs in naive BmN4 cells (right) or type-E pre-piRNAs in Tri-KO cells (left), separately analysed for Siwi-dominant (top) and BmAgo3-dominant (bottom) piRNA loci. The 6 nucleotides in the

BmZuc motif are highlighted. **g**, Schematic explanation for the weighted BmZuc motif (top) and the calculation of the BmZuc score in the 17-nt sliding window analysis (bottom). **h**, Similarity scores with the weighted BmZuc motif (BmZuc score) for Siwi were calculated for 111750 RNA and their mutant sequences in sliding windows and plotted as in **c**. **i**, Box plots show the maximum similarity scores with the weighted BmZuc motif for Siwi or BmAgo3 within the positions of 19–45 nt of Siwi-dominant (top) or BmAgo3-dominant (bottom) type-N or type-E piRNA loci or the shuffled control sequences (a pool of 3,236 species of 27-nt scrambled sequences that have the average nucleotide composition of the silkworm genome). Type-E piRNA loci have significantly higher BmZuc scores than the shuffled control sequences for both Siwi- and BmAgo3-dominant piRNAs (Mann–Whitney *U* test). Centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers.



Extended Data Fig. 6 | Nucleotide preference around the cleavage site by mouse MitoPLD. **a**, Peak length distribution of 2'-O-methylated small RNAs in wild-type (WT) or *Pnlcd1*^{-/-} mouse secondary spermatocytes. Data are from ref. ². **b**, Nucleotide composition around the 3' end of small RNAs in WT (left) or

Pnlcd1^{-/-} mice. The 3' ends of pre-piRNAs in *Pnlcd1*^{-/-} mice exhibit strong +1U bias. In addition, -9A and -3C preferences, which are similar to the BmZuc motif (Fig. 4e), are observed in fully elongated 37–42 nt pre-piRNAs in *Pnlcd1*^{-/-} mice (right). Data are from ref. ².

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
 - ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
 - ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
 - ☐ ☒ A description of all covariates tested
 - ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
 - ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
 - ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
 - ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
 - ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
 - ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection FV31S-SW (version 2.1.1.98), Multi Gauge (version 3.0)
Adobe Photoshop Elements (version 10)

Data analysis Data analysis was performed using cutadapt (version 1.16), Bowtie (version 1.2.2), bedtools (version 2.27.1), samtools (version 1.3), R version 3.4.0. Packages used include: Biostrings (version 1.1.2), coin (version 1.3.0), ggplot2 (version 2.2.1), gridExtra (version 2.3), lsr (version 0.5), RcolorBrewer (version 1.1.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The accession number for the sequencing data is DDBJ: DRA008549.

For Extended Data Fig. 6, publicly available datasets were used (SRR7760309, SRR7760310, SRR7760317, SRR7760318, SRR7760321, SRR7760322, SRR7760343, SRR7760344, SRR7760347, SRR7760348, SRR7760369, SRR7760370, SRR7760373, SRR7760374, SRR7760377, SRR7760378).

The code for bioinformatic analysis can be found at https://github.com/kshoji-nt/BmZuc_cleavage/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods was used to predetermine the sample size.
Data exclusions	No data exclusion in this manuscript.
Replication	The number of each experiment is described in Statistics and Reproducibility section in Method. All attempts at replication were successful.
Randomization	Randomization was not done, because this study does not involve animal experiments.
Blinding	Blinding is not relevant to this study since this is an exploratory study and blinding is impossible or unlikely to affect the results or interpretation of the results.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Randomization	If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.
Did the study involve field work?	<input type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access and import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<p>[Primary antibodies]</p> <p>Anti-Siwi, BmAgo3, BmZuc, BmArmi, BmGPAT1, BmGasZ, Trimmer, BmPapi antibodies were generated by immunizing rabbits. Anti-FLAG (M2, Sigma #F1804), Actin (Santa Cruz #sc-1616), GFP (B-2, Sabta Cruz#sc-9996) antibodies were purchased. For Westernblotting, each antibody was used at the following dilution [anti-Siwi (1:2000), BmAgo3 (1:2000), BmZuc (1:200), BmArmi (1:1000), BmGPAT1 (1:2000), BmGasZ (1:2000), Trimmer (1:1000), BmPapi (1:1000), FLAG (1:2000), Actin (1:2000), GFP (1:2000)]. For immunofluorescence, anti-BmGasZ antibody was used at 1:400.</p> <p>[Secondary antibodies]</p> <p>Peroxidase AffiniPure F(ab')₂ Fragment Goat Anti-Rabbit IgG, Fc fragment specific (Jackson ImmunoResearch #111-036-046 ; 1:5000)</p> <p>Peroxidase AffiniPure Goat Anti-Mouse IgG, light chain specific (Jackson ImmunoResearch #115-035-174 ; 1:5000)</p> <p>Peroxidase AffiniPure Donkey Anti-Goat IgG (H+L) (Jackson ImmunoResearch #705-035-003 ; 1:5000)</p> <p>Alexa Fluor 647 donkey anti-rabbit IgG antibody (ThermoFisher/Invitrogen # A-31573; 1:4000)</p>
Validation	[Primary antibodies]

Anti-Siwi, BmAgo3, BmZuc, BmArmi, BmGPAT1, BmGasZ antibodies were validated by Western blotting of a knockdown lysate of BmN4 (silkworm) in this paper (Fig. 3f). Anti-Trimmer, BmPapi antibodies were validated by Western blotting of a knockdown lysate of BmN4 (silkworm) in a previous study (2016 Cell 164, 962-973 Izumi et al.).
 For the following purchased antibodies, see manufacturer information:
 Anti-FLAG antibody (<https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin/f1804bul.pdf>)
 Anti-GFP antibody (<https://datasheets.scbt.com/sc-9996.pdf>)
 Anti-Actin antibody (https://search.cosmobio.co.jp/cosmo_search_p/search_gate2/docs/SCB_/SC1616.20070823.pdf)
 species: mouse, rat, human, zebrafish, C. elegans, Drosophila, S. cerevisiae and Xenopus
 applications: Western blotting, immunoprecipitation, immunofluorescence, and flow cytometry
 [Secondary antibodies]
 Peroxidase AffiniPure F(ab')₂ Fragment Goat Anti-Rabbit IgG, Fc fragment specific (<https://www.jacksonimmuno.com/catalog/products/111-036-046>)
 Peroxidase AffiniPure Goat Anti-Mouse IgG, light chain specific (<https://www.jacksonimmuno.com/catalog/products/115-035-174>)
 Peroxidase AffiniPure Donkey Anti-Goat IgG (H+L) (<https://www.jacksonimmuno.com/catalog/products/705-035-003>)
 Alexa Fluor 647 donkey anti-rabbit IgG antibody (<https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-31573>)

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s) BmN4 cells are provided from Dr. Kusakabe, Kyushu University

Authentication BmN4 cells have not been authenticated.

Mycoplasma contamination Not tested

Commonly misidentified lines (See [ICLAC](#) register) No commonly misidentified cell lines were used.

Palaeontology

Specimen provenance Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session

(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software	<i>Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.</i>
Cell population abundance	<i>Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.</i>
Gating strategy	<i>Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.</i>

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	<i>Indicate task or resting state; event-related or block design.</i>
Design specifications	<i>Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.</i>
Behavioral performance measures	<i>State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).</i>

Acquisition

Imaging type(s)	<i>Specify: functional, structural, diffusion, perfusion.</i>
Field strength	<i>Specify in Tesla</i>
Sequence & imaging parameters	<i>Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.</i>
Area of acquisition	<i>State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.</i>
Diffusion MRI	<input type="checkbox"/> Used <input type="checkbox"/> Not used

Preprocessing

Preprocessing software	<i>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</i>
Normalization	<i>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</i>
Normalization template	<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal	<i>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</i>
Volume censoring	<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>

Statistical modeling & inference

Model type and settings	<i>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</i>
Effect(s) tested	<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:	<input type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See Eklund et al. 2016)	<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
Correction	<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>

Models & analysis

n/a	Involvement in the study
<input type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.

Processive extrusion of polypeptide loops by a Hsp100 disaggregase

<https://doi.org/10.1038/s41586-020-1964-y>

Received: 17 May 2019

Accepted: 3 December 2019

Published online: 29 January 2020

There are amendments to this paper

Mario J. Avellaneda¹, Kamila B. Franke², Vanda Sunderlikova¹, Bernd Bukau², Axel Mogk² & Sander J. Tans^{1,3*}

The ability to reverse protein aggregation is vital to cells^{1,2}. Hsp100 disaggregases such as ClpB and Hsp104 are proposed to catalyse this reaction by translocating polypeptide loops through their central pore^{3,4}. This model of disaggregation is appealing, as it could explain how polypeptides entangled within aggregates can be extracted and subsequently refolded with the assistance of Hsp70^{4,5}. However, the model is also controversial, as the necessary motor activity has not been identified^{6–8} and recent findings indicate non-processive mechanisms such as entropic pulling or Brownian ratcheting^{9,10}. How loop formation would be accomplished is also obscure. Indeed, cryo-electron microscopy studies consistently show single polypeptide strands in the Hsp100 pore^{11,12}. Here, by following individual ClpB–substrate complexes in real time, we unambiguously demonstrate processive translocation of looped polypeptides. We integrate optical tweezers with fluorescent-particle tracking to show that ClpB translocates both arms of the loop simultaneously and switches to single-arm translocation when encountering obstacles. ClpB is notably powerful and rapid; it exerts forces of more than 50 pN at speeds of more than 500 residues per second in bursts of up to 28 residues. Remarkably, substrates refold while exiting the pore, analogous to co-translational folding. Our findings have implications for protein-processing phenomena including ubiquitin-mediated remodelling by Cdc48 (or its mammalian orthologue p97)¹³ and degradation by the 26S proteasome¹⁴.

We studied the disaggregase ClpB, a member of the Hsp100 chaperone family, using single-molecule techniques. Maltose-binding protein (MBP) was coupled to DNA handles at both termini and tethered between polystyrene beads, which were trapped and manipulated with optical tweezers (Fig. 1a). After mechanical unfolding of the protein (Fig. 1a, Extended Data Fig. 1a), the applied force was reduced to a value between 5 and 10 pN, high enough to prevent spontaneous refolding (Fig. 1a). Addition of ATP and ClpB(Y503D)—a mutant altered in the regulatory middle (M) domain that does not require Hsp70 (DnaK) binding for ATPase activation¹⁵—resulted in isolated episodes of contraction in the bead-to-bead distance (Fig. 1b). Zooming in showed that the effective polypeptide contour length L_e was initially approximately 360 amino acids (aa), as expected for fully unfolded MBP, and then decreased linearly to 0, indicating that the C and N termini of MBP were directly adjacent to each other (Fig. 1c, Extended Data Fig. 1b). After a brief pause, L_e increased abruptly back to 360 residues and then immediately decreased again (Fig. 1c). ClpB thus produced processive substrate translocation runs that ended with a loss of ClpB grip. This in turn caused the substrate to slip and be pulled back by the applied force and hence enabled a new run to start.

Translocation was abolished by using ADP instead of ATP; when either of the two ClpB ATPase catalytic centres (E279A or E678A) were mutated, preventing ATP hydrolysis; when the substrate-contacting pore loops (Y251A or Y653A) were mutated; or by deletion of the

N-terminal domain that forms the pore entry. Translocation was also observed for the M-domain mutant K476C and for wild-type ClpB with the Hsp70 (DnaK in bacteria) system¹⁶ (Extended Data Fig. 2a–e). ClpB(K476C) and wild-type ClpB translocated at the same speed as ClpB(Y503D), which was unexpected because both stimulated ATP hydrolysis less strongly in bulk¹⁵ (Extended Data Fig. 2f, g). However, they exhibited translocation for a smaller proportion of the time (Extended Data Fig. 2b), suggesting that the differences in hydrolysis rates reflected the fraction of actively translocating ClpB hexamers rather than their individual translocation speed.

Longer polypeptide constructs of two and four tandem MBP repeats displayed longer runs before slipping, with some exceeding 1,000 residues (Extended Data Figs. 1c–f, 3, 4). The speed distribution displayed two peaks (the second at roughly double the speed of the first) and extended beyond 500 aa per second (Fig. 1d), more than tenfold faster than other peptide translocases^{17–19}. This distribution appeared similar for the different substrate constructs and for different individual translocation bursts (Fig. 1b, Extended Data Fig. 3d–f). These bursts probably reflected the activity of single ClpB hexamers, because they consisted of continuous run–slip–run activity and were spaced apart by several seconds. Without slowing down, ClpB exerted high forces of more than 50 pN, resulting in the melting of our DNA tethers (Fig. 1e). These data indicated notable speed, processivity and power.

¹AMOLF, Amsterdam, The Netherlands. ²Center for Molecular Biology of Heidelberg University, German Cancer Research Center, Heidelberg, Germany. ³Department of Bionanoscience, Kavli Institute of Nanoscience Delft, Delft University of Technology, Delft, The Netherlands. *e-mail: s.tans@amolf.nl

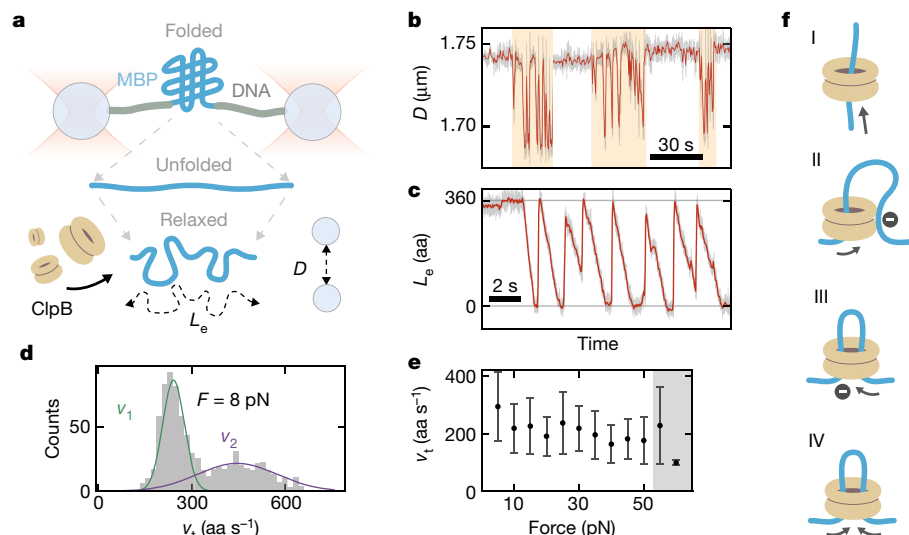


Fig. 1 | ClpB is a processive translocase. **a**, Tethered MBP was unfolded with optical tweezers, relaxed to a low force that prohibits refolding, and exposed to ClpB and ATP. **b**, Tether contraction bursts (orange regions) with ClpB (Y503D) and ATP. Grey, raw signal (500 Hz); red, filtered signal (2 Hz). **c**, Polypeptide contour length L_e during a contraction burst, as determined from the bead-bead distance (D), force, and worm-like chain model. L_e decreases linearly from 360 aa (MBP fully extended) down to 0 aa (MBP C and N termini directly adjacent), indicating processive translocation by ClpB. Abrupt increases of L_e to 360 aa indicate that ClpB transiently loses grip and substrate slips backwards, pulled by the applied force. Red, filtered signal (20 Hz). **d**, Speed

distribution of runs from all MBP substrate constructs in the presence of ClpB (Y503D) and ATP, at a force of approximately 8 pN. Double Gaussian fit shows two mean speeds, $v_2 \approx 2v_1$ with $v_1 = 240 \pm 30 \text{ aa s}^{-1}$ and $v_2 = 450 \pm 130 \text{ aa s}^{-1}$ (mean \pm s.d., $n = 800$ runs, 18 molecules). **e**, Mean translocation speed for ClpB (Y503D) versus applied tension ($n = 717$ runs, 8 molecules; see Methods). Grey, DNA melting regime and upper force limit. Data are mean \pm s.e.m. **f**, Hypothetical ClpB translocation topologies. Single-strand insertion and translocation (I) does not yield contraction, unless it is immobilized elsewhere on the ClpB surface (II). Dual-strand insertion in a looped topology can also produce contraction, either by single-arm (III) or dual-arm (IV) translocation.

Different hypothetical translocation models or topologies could be considered (Fig. 1f). Even when their termini are not free for insertion, as is the case here, single polypeptide chains can be accommodated into the ClpB pore by rings that open and close or that assemble around them^{20,21} (model I). However, this scenario would only produce the observed contractions if a second chain site is immobilized on ClpB (model II), analogous to DNA processing by condensin²². Alternatively, the substrate could be inserted as a loop into the central pore, with translocation of one (model III) or both (model IV) arms of the loop.

Testing these models with optical tweezers is difficult. We therefore developed a technique that allows independent measurement of the length of each arm of the polypeptide loop and integrates optical tweezers with ClpB tracking at sub-wavelength resolution using single-molecule fluorescence imaging (Fig. 2a, Extended Data Fig. 5). We chose the construct with two maltose-binding protein (MBP) repeats (2MBP), as it yields longer runs, and exposed it to fluorescently labelled ClpB and ATP after unfolding, while scanning a confocal excitation beam along the tether and beads (Fig. 2b). To limit the parasitic signal emanating from the beads, we developed a protein–DNA coupling protocol that enabled the attachment of long 5-kilobase pair (kbp) DNA handles (see Methods). Single ClpB-binding events were identified by a fluorescent spot appearing between the beads (Fig. 2b), and translocation was observed soon after (Fig. 2c). We next moved to an ATP-only solution to reduce background fluorescence and prevent further ClpB binding, and tracked the spot position using Gaussian fitting (Fig. 2d–f). Combining the tweezers and tracking data yielded the distances between ClpB and each of the MBP termini, and hence the translocation activity on both loop arms independently (Fig. 2a, Methods).

We found various sequences of events: after translocation of the entire chain, the left arm of the loop was released and slipped backwards until the full chain was again extended in *cis*, and subsequently left-arm translocation restarted rapidly (Fig. 2g, h, event sequence A \rightarrow B \rightarrow C). A similar sequence on the right side occurred directly afterwards

(Fig. 2g, h, A \rightarrow D \rightarrow E). We also observed both arms being translocated simultaneously, each at similar velocity (Fig. 2g, i, event F, Extended Data Fig. 5i–k). Consistently, the total translocation speed, which reflects the velocity at which both polypeptide termini approach each other, and is more accurate as it is based only on the signal from the optical tweezers, was then twice as high ($2v$) as in single-arm translocation runs (v) (Fig. 2k, grey region). Model II does not allow for two-arm translocation and hence was not consistent with the data, whereas models III and IV were consistent with the data. Switches between single- and two-arm translocation modes took place after blockage of one arm, typically on ClpB encountering the DNA tether at either terminus. The data also provided direct confirmation that single ClpB rings remained intact and bound during runs, switches and back-slips (Fig. 2h, i).

This scenario was supported by multiple additional observations. First, 64% of the very first runs in a translocation burst initially showed the higher speed ($2v$) before switching to the lower speed (v), compared with 22% when considering all runs (Extended Data Fig. 6a). Indeed, the initial ClpB binding site is probably not directly adjacent to the DNA handles at the termini, and thus both arms are then unobstructed when translocation starts. Initial binding regions estimated from these experiments were consistent with peptide scanning data, although we note that both methods yield rough estimates (Extended Data Fig. 6). Second, experiments at increased resolution showed that lower-speed (v) runs were composed of individual translocation steps of $14.6 \pm 0.9 \text{ aa}$, whereas higher-speed ($2v$) runs were in steps of $28 \pm 3 \text{ aa}$ (Fig. 3a–d, Extended Data Fig. 7a–e). These findings are consistent, since decreases in distance between termini should be twofold larger when both arms are translocated simultaneously. ClpB thus switches between translocation modes by changing the step size rather than the step frequency (Fig. 3e).

We next investigated how these stepping dynamics relate to the structure of ClpB. Each ClpB monomer is thought to move substrates by approximately 2 aa, substantially less than the observed 14-aa or 28-aa steps^{12,23}. A proposed concerted action²⁴ of all subunits together

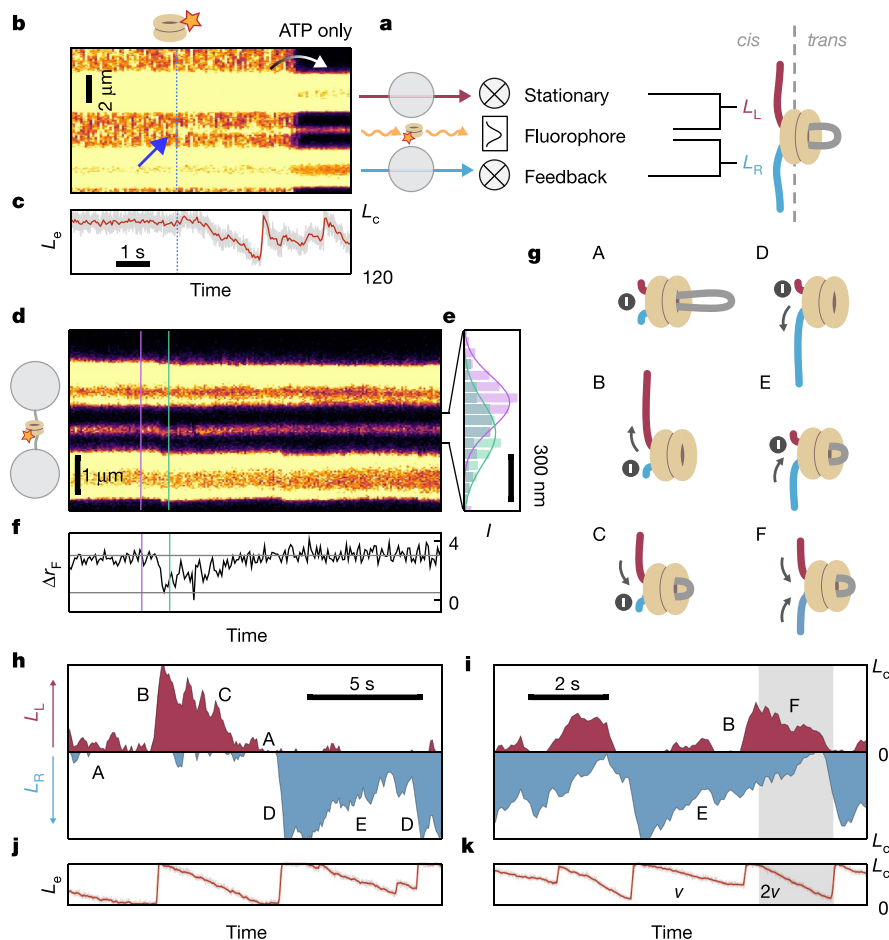


Fig. 2 | Optical tweezers with fluorescence reveals ClpB translocation of both loop arms. **a**, Principle of approach: one trap is continuously moved to maintain force constant. Bead positions yield polypeptide N-to-C-terminal distance at nanometre precision (expressed in contour length L_c). Confocal fluorescence imaging of ClpB-Atto633 yields its position at sub-wavelength precision using Gaussian fitting. Together, they quantify the lengths of both non-translocated (*cis*) polypeptide arms: L_R (blue) and L_L (purple). **b**, Fluorescence kymograph from scans along beads and tether, showing ClpB binding (blue arrow) and movement to the ClpB-free region. **c**, Concurrent tweezers data of polypeptide contour length L_c , showing translocation start soon after ClpB binding. **d**, Kymograph during translocation. **e**, Photon count of ClpB spot along two scans and Gaussian fits that determine position. **f**, Position of ClpB-Atto633 (in number of pixels), moving suddenly down at back-slip (**h**) and gradually up during translocation. **g**, Back-slip D does not change ClpB position, because slip is on the right (blue), and left-arm linked to stationary bead remains unchanged (purple). Top line, ClpB at left-hand MBP terminus; bottom line: ClpB is at right-hand terminus and polypeptide is fully in *cis*. Consistently, ClpB deviates from top line when tweezers detects back-slip. **h**, **i**, L_R and L_L for kymographs shown in **d** and Extended Data Fig. 5g. Grey-shaded region, double-speed translocation. Consistently, both arms shorten simultaneously. **j**, **k**, Total *cis*-polypeptide length, $L_c = L_R + L_L$ from tweezers alone.

would yield a continuous series of 2-aa steps, and thus appears inconsistent with these results. By contrast, the six subunits acting in rapid consecutive manner would produce steps similar to those detected here (approximately 12 aa, or approximately 24 aa when both arms are moving). The pauses between steps could thus reflect a slow transition within the ATP cycle occurring in all subunits²⁵. Zooming into the steps should show six substeps of around 2 aa, but these cannot be resolved owing to the particularly high translocation speed and the need to time-average at these length scales. Therefore, we mixed ATP with the poorly hydrolysable ATP γ S, as this would be expected to interrupt sequential

subunit action moving along the ClpB ring and therefore yield smaller steps. Translocation was more erratic and indeed showed smaller steps well below 14 aa in size (Extended Data Fig. 7g–i), rather than longer pauses only. These data thus supported the sequence-pause model.

The complex dynamics observed thus far can be further complicated when folded structures are present within the looped polypeptide. Specifically, we found back-slips for 2MBP that were incomplete, with a segment of about 270 aa remaining on the *trans* side of ClpB (Fig. 4a, b). This is exactly the length of one MBP core, suggesting that the polypeptide folded after translocation (in line with its normal folding time of about 1 s)²⁶ and was subsequently blocked at the *trans* side of the narrow ClpB pore when pulled backwards during a back-slip (Fig. 4d). Consistently, such incomplete back-slips only occurred after full MBP cores were translocated (Fig. 4c) and were not observed for folding-compromised 2MBP mutants²⁷ or for 1MBP, whose core cannot fold because a key segment remains stuck in the ClpB pore (Extended Data Fig. 8). Of note, substrates refolded at the exit of the ClpB channel, analogous to co-translational folding of nascent chains, and without requiring DnaK.

Conversely, misfolded structures already present within the chain should be blocked at the *cis* side of ClpB during translocation. Such an obstruction of translocation can ramp up local forces and in turn pull apart the blocking misfolded structure. Indeed, we observed such disruption when a partially folded MBP or a small MBP aggregate was exposed to ClpB (Extended Data Fig. 9). The disruption events were directly followed by translocation runs because unfolded polypeptides produced by structure disruption on the *cis* side of ClpB are available for translocation. These data indicate how folded structures present in *cis* and *trans* can affect translocation dynamics in a looped topology.

In conclusion, our study on ClpB shows unambiguously that polypeptide loop extrusion is possible. Free substrate termini may also

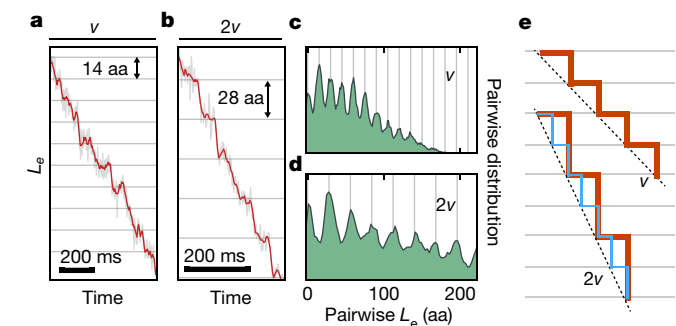


Fig. 3 | Translocation steps by ClpB. **a**, **b**, Plot of L_c for single-speed (**a**) and dual-speed (**b**) translocation runs. Red, Savitzky–Golay filtering. **c**, **d**, Distribution of L_c difference between any two points for one single-speed run (**c**) or one dual-speed run (**d**). The regularly spaced peaks indicate a step size of 14.6 ± 0.9 aa for the single-speed run; the peak spacing for the dual-speed run is doubled, yielding a step-size of 28 ± 3 aa. Data are mean \pm s.e.m. calculated from $n = 12$ runs. **e**, The data show that the speed is doubled by doubling the step size (red), not the step frequency (blue).

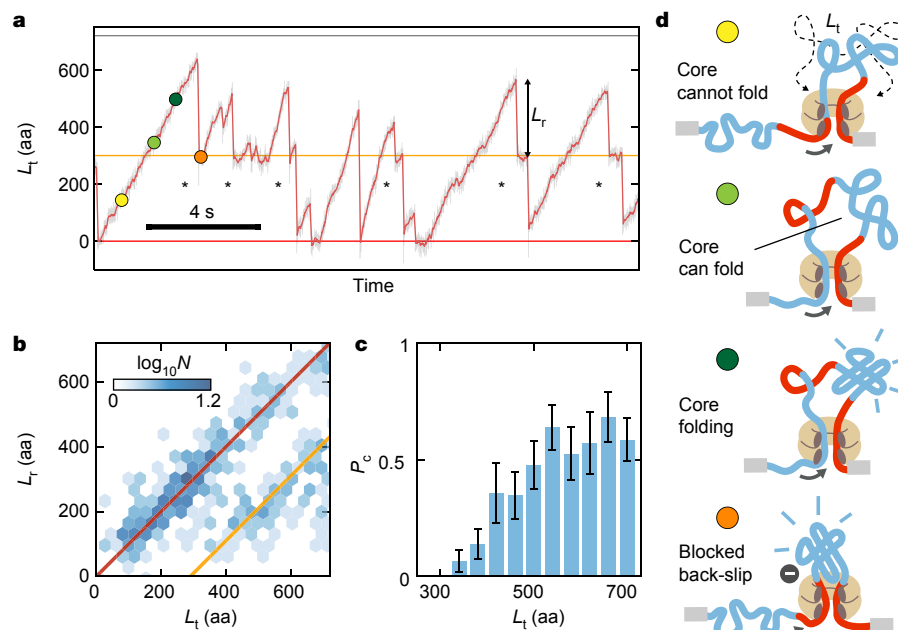


Fig. 4 | Substrate refolding on the ClpB *trans*-side during translocation. **a**, Translocated polypeptide length (L_t) for 2MBP. Horizontal red and orange lines, $L_t = 0$ and $L_t = 310$ aa, respectively (a 270-aa MBP core in *trans*, and 2×20 aa inside the ClpB pore). Stars, back-slip arrested at $L_t = 310$ aa. Circles, events illustrated in **d**. **b**, Red line, translocated length is fully released to the *cis* side upon back-slip ($L_r = L_t$). Orange line corresponds to an MBP core refolding on the *trans* side and then blocked from passing the ClpB pore upon back-slip ($L_r = L_t - 310$ aa). **c**, Probability of core folding (P_c). Consistently, P_c is 0 below 310 aa, when the MBP core is not fully translocated ($n = 266$ runs, 12 molecules, mean \pm standard error of binomial distribution, see Methods). **d**, Cartoons of *trans* refolding. Blue, core MBP; red, non-core MBP. MBP cores do not refold on the *cis* side because tweezers constrain the polypeptide chain. In cells, aggregates may constrain substrates from folding. On the *trans* side, the chain is freed from these constraints and an MBP core can fold when fully translocated (light and dark green dots) and thus cannot pass through the ClpB pore when back-slip occurs (orange dot and line).

insert into the ClpB pore and be translocated in a non-loop topology, though we surmise that internal segments of aggregated proteins are targeted more readily and hence translocated as loops. ClpB is fast, processive, generates large forces, and can switch between single- and dual-arm translocation. ClpB thus appears to maintain a tight and long-term grip on both arms, with back-slips indicating a sporadic loss of contact. It remains an open question how the independent handling of two arms is achieved at the structural level. These features of ClpB are relevant to efficient disaggregation (Extended Data Fig. 10). Full dissolution of stable aggregates probably involves multiple ClpB rings and other chaperones such as Hsp70/DnaK acting at different sites, at different moments in time, and involving many random dissociation and re-association events. Nevertheless, ClpB translocation itself is remarkably deterministic and processive once started.

Overall, our findings define loop extrusion as the mechanistic basis of Hsp100 disaggregation, highlight the need for tight regulation of Hsp100 activity and suggest that other polypeptide processing systems such as the Cdc48 (mammalian orthologue p97) segregase, the ribosomal assembly factor Rix7 (mammalian orthologue NVL), and the 26S proteasome may also exploit the capability to handle multiple polypeptide strands in a controlled manner.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1964-y>.

- Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1332–1340 (2018).
- Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
- Weibezahn, J. et al. Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **119**, 653–665 (2004).
- Mogk, A., Bukau, B. & Kampina, H. H. Cellular handling of protein aggregates by disaggregation machines. *Mol. Cell* **69**, 214–226 (2018).
- Kummer, E. et al. Bacterial and yeast AAA+ disaggregases ClpB and Hsp104 operate through conserved mechanism involving cooperation with Hsp70. *J. Mol. Biol.* **428**, 4378–4391 (2016).

- Shorter, J. & Southworth, D. R. Spiraling in control: structures and mechanisms of the Hsp104 disaggregase. *Cold Spring Harb. Perspect. Biol.* **11**, 034033 (2019).
- Sousa, R. Structural mechanisms of chaperone mediated protein disaggregation. *Front. Mol. Biosci.* **1**, 12 (2014).
- Liberek, K., Lewandowska, A. & Ziętkiewicz, S. Chaperones in control of protein disaggregation. *EMBO J.* **27**, 328–335 (2008).
- Li, T. et al. *Escherichia coli* ClpB is a non-processive polypeptide translocase. *Biochem. J.* **470**, 39–52 (2015).
- Durie, C. L. et al. Hsp104 and potentiated variants can operate as distinct nonprocessive translocases. *Biophys. J.* **116**, 1856–1872 (2019).
- Gates, S. N. et al. Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **357**, 273–279 (2017).
- Deville, C., Franke, K., Mogk, A., Bukau, B. & Saibil, H. R. Two-step activation mechanism of the ClpB disaggregase for sequential substrate threading by the main ATPase motor. *Cell Rep.* **27**, 3433–3446 (2019).
- Bodnar, N. O. et al. Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1-Npl4. *Nat. Struct. Mol. Biol.* **25**, 616–622 (2018).
- Dong, Y. et al. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49–55 (2019).
- Oguchi, Y. et al. A tightly regulated molecular toggle controls AAA+ disaggregase. *Nat. Struct. Mol. Biol.* **19**, 1338–1346 (2012).
- Fernández-Higuero, J. A., Aguado, A., Perales-Calvo, J., Moro, F. & Muga, A. Activation of the DnaK–ClpB complex is regulated by the properties of the bound substrate. *Sci. Rep.* **8**, 5796 (2018).
- Maillard, R. A. et al. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459–469 (2011).
- Aubin-Tam, M.-E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257–267 (2011).
- Olivares, A. O., Nager, A. R., Iosefson, O., Sauer, R. T. & Baker, T. A. Mechanochemical basis of protein degradation by a double-ring AAA+ machine. *Nat. Struct. Mol. Biol.* **21**, 871–875 (2014).
- Werbeck, N. D., Schlee, S. & Reinstein, J. Coupling and dynamics of subunits in the hexameric AAA+ chaperone ClpB. *J. Mol. Biol.* **378**, 178–190 (2008).
- Uchihashi, T. et al. Dynamic structural states of ClpB involved in its disaggregation function. *Nat. Commun.* **9**, 2147 (2018).
- Ganji, M. et al. Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
- Yu, H. et al. ATP hydrolysis-coupled peptide translocation mechanism of *Mycobacterium tuberculosis* ClpB. *Proc. Natl Acad. Sci. USA* **115**, E9560–E9569 (2018).
- Lyubimov, A. Y., Strycharzka, M. & Berger, J. M. The nuts and bolts of ring-translocase structure and mechanism. *Curr. Opin. Struct. Biol.* **21**, 240–248 (2011).
- Moffitt, J. R. et al. Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446–450 (2009).
- Bechtluft, P. et al. Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**, 1458–1461 (2007).
- Chakraborty, K. et al. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**, 112–122 (2010).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification

All MBP constructs were modified at both termini with cysteine residues using the pET28 vector. Double-mutant MBP harbours two mutations (V8G and Y283D) that hinder folding²⁸. Proteins were purified from *Escherichia coli* BL21(DE3) cells. For overexpression, overnight cultures were diluted 1:100 in fresh LB medium supplemented with 50 mg l⁻¹ kanamycin, 0.2% glucose and incubated under vigorous shaking at 30 °C. Expression was induced at OD₆₀₀ = 0.6 by addition of 1 mM IPTG and incubation overnight at room temperature. Cells were cooled, collected by centrifugation at 5000g for 20 min, flash-frozen and stored at -80 °C. Cell pellets were resuspended in ice-cold buffer A (50 mM potassium phosphate pH 7.5, 150 mM NaCl, 3 mM chloramphenicol, 50 mM Glu-Arg, 10 mM Complete Protease Inhibitor Ultra (Roche), 10 mM EDTA) and lysed using a pressure homogenizer. The lysate was cleared from cell debris by centrifugation at 50,000g for 60 min and incubated with Amylose resin (New England Biolabs) that was previously equilibrated in buffer A for 20 min at 4 °C. The resin was washed with buffer A three times by centrifugation and bound proteins were eluted in buffer A supplemented with 20 mM maltose. Purified proteins were aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C. ClpB and variants were overexpressed in *E. coli* ΔclpB::kan cells. Cell pellets were resuspended in LEW buffer (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol) and lysed by French press. Cleared supernatants were incubated with Protino Ni-IDA resin and bound proteins were eluted by LEW buffer containing 250 mM imidazole. ClpB containing fractions were pooled and subjected to Superdex S200 16/60 size-exclusion chromatography in MDH buffer (50 mM Tris pH 7.5, 150 mM KCl, 20 mM MgCl₂, 2 mM DTT) containing 5% (v/v) glycerol.

ClpB-Atto633 labelling

Labelling of ClpB-E731C variants with Atto633-maleimide was performed in PBS buffer according to the instructions of the manufacturer (ATTO-TEC). Labelled ClpB-E731C was separated from non-reacted Atto633 by size-exclusion chromatography using Superdex S200 HR10/30 in MDH buffer containing 5% (v/v) glycerol.

Attachment of DNA handles to substrates

Protein substrates were buffer-exchanged using a PD-10 desalting column (GE Healthcare) to remove reducing agents and elutants. Next, they were incubated with a 4× excess maleimide-modified oligonucleotides (20 nucleotides) for 1 h at 30 °C. Uncoupled oligos were removed using Amylose resin (NEB). The coupled protein was then ligated to 2.5-kbp DNA tethers presenting a complementary 20-nucleotide single-stranded overhang using T4 ligase for 1 h at room temperature.

Optical tweezers assay

Carboxyl polystyrene beads (CP-20-10, diameter 2.1 μm, Sphero-tech) were covalently coated with sheep anti-digoxigenin antibody (Roche) via a carbodiimide reaction (PolyLink Protein coupling kit, Polysciences). Approximately 50 ng of the generated construct were incubated with 2 μl beads in 10 μl buffer C (50 mM HEPES pH 7.5, 5 mM MgCl₂, 100 mM KCl) for 15 min in a rotary mixer at 4 °C and rediluted in 350 μl buffer C. With our coupling strategy, approximately 50% of the constructs were asymmetrically functionalized with digoxigenin and biotin in each side. In order to create the second connection, we used NeutrAvidin-coated polystyrene beads (NVP-20-5, diameter 2.1 μm, Sphero-tech). Once trapped, beads were brought into close proximity to allow binding, and tether formation was identified by an increase in force when the beads were moved apart. ClpB was diluted in buffer C

to a final concentration of 2 μM. For the ATP experiments, we used an ATP regeneration system (3 mM ATP, 20 ng μl⁻¹ pyruvate kinase, 3 mM phosphoenol pyruvate). Experiments were performed in the presence of an oxygen scavenging system²⁹ (3 units per ml pyranose oxidase, 90 units per ml catalase and 50 mM glucose, all purchased from Sigma-Aldrich) to prevent DNA and protein oxidation damage.

Single-molecule data analysis and ClpB translocation event characterization

Data was recorded at 500 Hz using a custom-built dual trap optical tweezers and a C-Trap (Lumicks). Data was analysed using custom scripts in Python. The optical traps were calibrated using the power spectrum of the Brownian motion of the trapped beads²⁸, obtaining average stiffness values of $\kappa = 0.39 \pm 0.04$ pN nm⁻¹. Most measurements were taken in an active force-clamp regime, in which one of the traps was moved in response to changes in the force using a proportional–integral–derivative (PID) feedback loop (Extended Data Fig. 5e, f). Individual force-extension curves were identified and fitted to two worm-like-chain (WLC) models in series (Extended Data Fig. 1a), using the approximation of an extensible polymer reported in ref.³⁰ for the DNA, and the Odijk inextensible approximation for the protein contribution³¹.

$$x = L_c^* \left[\frac{4}{3} \left(1 - \frac{1}{\sqrt{\beta^* + 1}} \right) - \frac{10e^{\frac{4\sqrt{900}}{\beta^*}}}{\sqrt{\beta^*} \left(e^{\frac{4\sqrt{900}}{\beta^*}} - 1 \right)^2} + \frac{\beta^{*1.62}}{3.55 + 3.8\beta^{*2.2}} \right] + L_e \left(1 - \frac{1}{2\sqrt{\beta}} \right)$$

Where $\beta^* = \frac{FL_p^*}{k_B T}$, F is the force, T is the temperature and L_p^* , K and L_c^* are the persistence length, stretch modulus and contour length of DNA, respectively; $\beta = \frac{FL_p}{k_B T}$, where L_p and L_e are the persistence and extended length of the protein, respectively. L_c^* values were 906, 1,750 and 3,500 nm for the three different DNA handles used (1.3, 2.5 and 5 kb, respectively), and L_p was 0.75 nm. L_p^* and K were fitted, yielding average values of 30 nm and 700 pN nm⁻¹, respectively. These fitted parameters were then used to compute the instantaneous extended length of the protein (L_e) using the same WLC model (Extended Data Fig. 1b). The translocated length (L_t) was computed by subtracting the extended length (L_e) to the total contour length of the protein (L_c). The unfiltered data (500 Hz) is displayed in all panels in grey. With the exception of Fig. 1b, the red signal always indicates data decimated to 20 Hz.

To classify translocation events, the translocated length signal was smoothed using a Savitzky–Golay filter³² (Extended Data Fig. 4c, black line), enabling its time derivative to be calculated without large fluctuations. Back-slipping results in a large negative slope in the derivative, which was used as the criteria to separate individual translocation runs (Extended Data Fig. 4d). Subsequent one-dimensional dilation and erosion was performed to remove artefacts. Next, each individual run was similarly treated in order to find local changes in the slope (Extended Data Fig. 4e), setting as threshold a value between the two known speeds (around 70 and 140 nm s⁻¹, Extended Data Fig. 4f). Linear fits were performed in each identified region and reported as the local translocation speed (Extended Data Fig. 4e). Only fits that yielded r values higher than 0.8 were considered unless otherwise stated. Speed distributions were computed using the speeds of all valid runs for each condition.

Translocation-step characterization

To increase the spatial resolution³³, we tethered a single MBP using 1,300-bp DNA handles, 500 μM ATP and high tension (>20 pN). Raw data was smoothed using a Savitzky–Golay filter of 5th order with a window of 21 data points. The difference between every distinct pair of data points was calculated and the sample was binned to compute

the pairwise length distribution. To calculate the periodicity more accurately, we computed the autocorrelation of the pairwise length distribution using the Pearson correlation coefficient for different lag lengths (Extended Data Fig. 7a, b) and its power spectrum using the Welch method³⁴ (Extended Data Fig. 7c, d). The autocorrelation distribution was fitted using the equation:

$$\left(A \cos\left(\frac{2\pi}{T}x\right) + mx + n \right) Be^{-Cx}$$

where T is the period of the steps and a linear and exponential function have been introduced to account for the decay in the signal (Extended Data Fig. 7a, b). The peak in the power spectrum was fitted to a Gaussian distribution (Extended Data Fig. 7c, d).

Confocal fluorescence measurements

An excitation laser beam with a wavelength of 638 nm and a power of 1.3 mW was scanned along the beads and tether at a line rate of 12 Hz. To avoid parasitic noise from the beads, proteins were tethered using two 5-kbp instead of 2.5-kbp DNA handles. In addition, the 2MBP construct was used in order to observe larger distance changes. Force spectroscopy and confocal microscopy data were synchronized based on the movement of the beads (Extended Data Fig. 5a–d). The edge of the moving bead was tracked using a Gaussian fit (Extended Data Fig. 5b) and overlying it on top of the optical tweezers signal for the bead movement showed a time offset (Extended Data Fig. 5c). In order to determine the value of this offset, we computed the root-mean-square deviation (r.m.s.d.) between the signals for different time offsets (Extended Data Fig. 5d):

$$\text{r.m.s.d.}(\tau) = \sqrt{\frac{\sum_i^{N(\tau)} (X_i(\tau) - x_i(\tau))^2}{N(\tau)}}$$

Where τ is the time offset applied to the tracked signal, $N(\tau)$ is the total number of points, $X(\tau)$ is the position of the bead according to the voltage of the mirror and $x(\tau)$ is the tracked position from the fluorescence kymograph. Minimization of r.m.s.d. (τ) provides an excellent estimate of the time offset between the signals (Extended Data Fig. 5d).

Integration of optical tweezers and imaging signals to compute the length components

After ClpB binding and moving to a region containing only ATP, the fluorescent spot between the beads was fitted to a Gaussian distribution. To reduce the noise of the signal, we averaged the intensity profiles of three scanning lines before fitting. The resulting trajectory yielded the absolute position of ClpB with subpixel precision, which was then converted to nanometres using a factor of 80 nm per pixel.

Next, we computed the position of each bead edge that is closest to ClpB (bottom edge for top bead and vice versa) using the trap position, bead displacement and bead radius. Although it is possible to obtain these positions from the fluorescence kymograph, the optical tweezers data yield higher spatial resolution. We subtracted the ClpB tracked position from the position of bottom edge of the top bead, and we subtracted the position of the top edge of the bottom bead from the ClpB position. These distances contain an arbitrary shift owing to the mismatched reference system of the optical tweezers and confocal fluorescence images. In order to identify the offset, we used the fact that when the polypeptide is completely translocated (information present in the optical tweezers data, such as Fig. 1c or Extended Data Fig. 2), both distances should be equal to each other and equal to half the distance D between the edges of the beads. After correcting for the shift, we obtained the absolute distance between ClpB and each of the beads (D_L and D_R). Since we use a force clamp, any change in distance is solely due to a change in the protein length ($\Delta x_L = \Delta D_L$ and $\Delta x_R = \Delta D_R$, Extended Data Fig. 5e, f). Therefore, we removed the constant DNA

contribution and computed the protein contour length from each distance (L_L and L_R) using the WLC model.

Peptide library data and initial ClpB binding location

The MBP peptide library was prepared by automated spot synthesis by JPT Peptide Technologies (PepSpots). The library is composed of 13-mer peptides scanning the MBP primary sequence with an overlap of 10 residues. One micromolar ClpB-NTD (Met1–Ser148) was incubated for 30 min in buffer P (10 mM Tris pH 7.5, 150 mM KCl, 20 mM MgCl₂, 5% (w/v) sucrose and 0.005% (v/v) Tween 20) with the library. Afterwards, buffer P was discarded and the membrane was washed with cold TBS (50 mM Tris pH 7.6, 150 mM NaCl). Fractionated western blotting enabled transfer of ClpB-NTD bound to peptide spots onto PVDF membranes and bound ClpB-NTD was detected by use of specific, polyclonal (rabbit) anti-ClpB-NTD serum.

The obtained blot image was divided in regions and the individual intensities were computed (Extended Data Fig. 6d). A Gaussian filter was applied to the resulting distribution to account for sequence overlapping and mirroring was performed for direct comparison with the optical tweezers data (Extended Data Fig. 6e).

ATPase activity assay

MBP–DM was denatured in 50 mM Tris, 25 mM KCl, 10 mM MgCl₂, 6 M urea and 2 mM DTT. The ATPase activity of the different variants was determined in 50 mM Tris, 25 mM KCl, 10 mM MgCl₂, 0.4 M urea and 2 mM DTT in presence of 2 mM ATP. ATPase measurements were started by addition of substrate.

Additional statistical calculations

Error bars of proportion histograms (Fig. 4c and Extended Data Fig. 6a) were calculated using the standard error of a binomial distribution:

$$\sigma = \sqrt{\frac{p(1-p)}{N}}$$

where p is the success proportion and N is the total number of observations.

Statistical sizes for bar plots are: Fig. 1e: 20, 9, 52, 58, 31, 48, 77, 40, 4, 41, 14 and 5 for each point; Fig. 4c: 41, 30, 29, 14, 23, 23, 25, 19, 14, 19 and 29 for each bar.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

Code availability

All data were analysed using a custom Python package that is available online and can be downloaded upon request to the corresponding author.

- Berg-Sørensen, K. & Flyvbjerg, H. Power spectrum analysis for optical tweezers. *Rev. Sci. Instrum.* **75**, 594–612 (2004).
- Swoboda, M. et al. Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS Nano* **6**, 6364–6369 (2012).
- Petrosyan, R. Improved approximations for some polymer extension models. *Rheol. Acta* **56**, 21–26 (2017).
- Odijk, T. Stiff chains and filaments under tension. *Macromolecules* **28**, 7016–7018 (1995).
- Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
- Forns, N. et al. Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles. *Biophys. J.* **100**, 1765–1774 (2011).

34. Welch, P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–73 (1967).

Acknowledgements We thank E. Koers for help with the substrate constructs and M. E. Aubin-Tam, K. Ganzinger, S. Werner and F. Wruck for comments and critical reading of the manuscript. This study was supported by the Netherlands Organization for Scientific Research (NWO) and by grants of the Deutsche Forschungsgemeinschaft (BB617/17-2, MO970/4-2 and MO 970/4-3) to B.B. and A.M., and the AmPro program of the Helmholtz Society to B.B.

Author contributions M.J.A., B.B., A.M. and S.J.T. conceived the research. K.B.F. and A.M. purified all chaperone variants and performed the biochemical assays. M.J.A. and S.J.T. designed the single-molecule experiments. M.J.A. and V.S. generated the substrate

constructs. M.J.A. carried out the single-molecule experiments, developed the Python software and performed the data analysis. M.J.A. and S.J.T. wrote the manuscript with the input and discussion of all authors.

Competing interests The authors declare no competing interests.

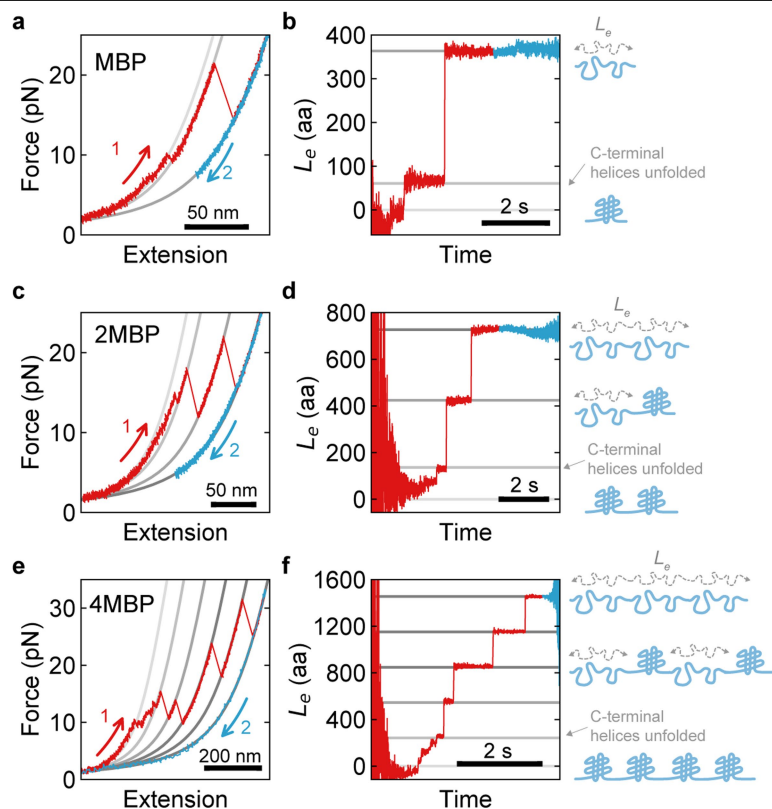
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1964-y>.

Correspondence and requests for materials should be addressed to S.J.T.

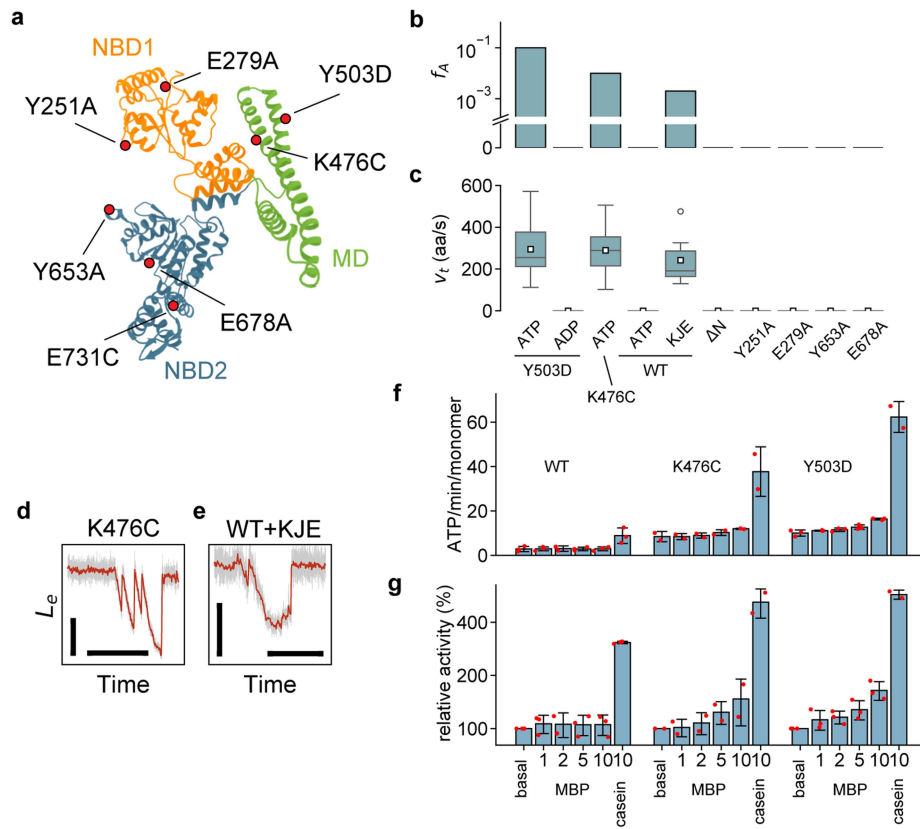
Peer review information *Nature* thanks Toshio Ando, Eilika Weber-Ban, Gijs Wuite and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



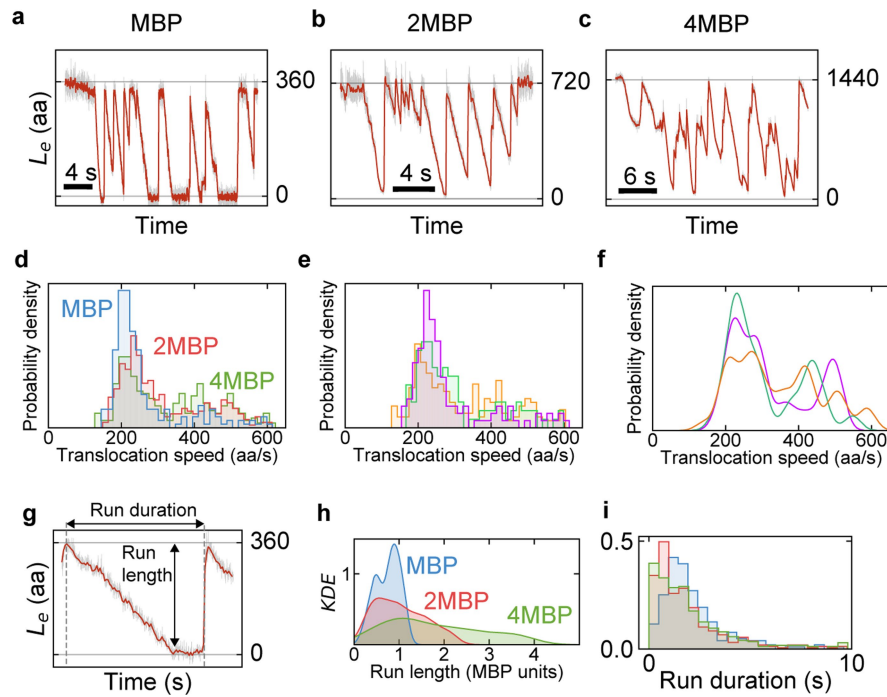
Extended Data Fig. 1 | Mechanical unfolding of substrates and extended length description. **a, c, e,** Force-extension curves showing the characteristic unfolding pattern: MBP (**a**), the 2MBP (**c**) and the 4MBP construct (**e**), with an initial gradual and discrete unfolding of C-terminal α -helices (Extended Data Fig. 8a) followed by a sharp unfolding of the cores. Grey lines show WLC fits to the data. Red indicates pulling and blue indicates relaxing of the protein chain. **b, d, f,** The corresponding extended length L_e of MBP (**b**), the 2MBP (**d**) and the

4MBP construct (**f**). L_e reflects the contour length along the polypeptide backbone, but only of the unfolded part of the protein that is compliant (that is, unfolded and at the *cis*-side of ClpB). L_e is determined from the measured force and extension (distance between beads), and using the WLC model of a non-interacting chain. Grey lines, contour length values obtained from the WLC fits. At low forces, the WLC curves of different contour lengths converge, yielding noisy data.



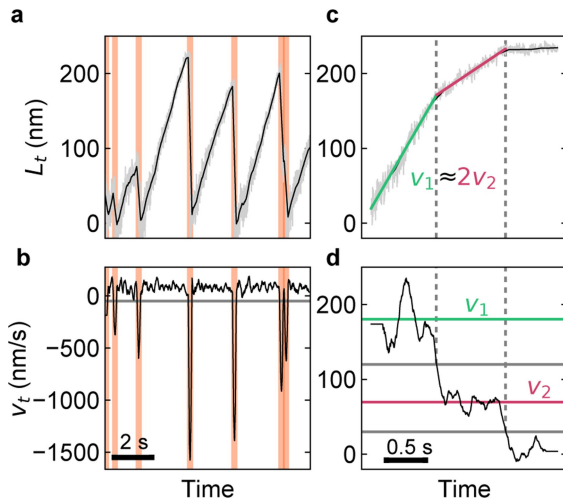
Extended Data Fig. 2 | Translocation by ClpB variants. **a**, ClpB monomer structure indicating all tested variants. These variants (except K476C) were generated in the constitutively active Y503D background. Variants E279A and E678A are Walker B mutants in the nucleotide-binding domains NBD1 and NBD2, respectively. These mutations abolish ATP hydrolysis at NBD1 or NBD2. Variants Y251A and Y653A are pore-loop mutants in NBD1 and NBD2, respectively. These mutations affect substrate interaction in the ClpB pore at either NBD1 or NBD2. The K476C variant undocks the middle domain (MD), mimicking the effect of Hsp70 (DnaK) activation. MD undocking in the Y503D variant is more pronounced, and therefore activation is more robust. An additional construct (ClpB(Δ N)) lacked the N-terminal domain (NTD), hindering initial substrate binding. Finally, the variant E731C harbours a cysteine at the bottom of NBD2 for fluorophore labelling. **b**, Fraction of time showing activity (f_A) for different mutants (in Y503D background, except K476C and wild type (WT)). **c**, Average translocation speed for all ClpB variants tested. KJE is the DnaK system (DnaK, DnaJ and GrpE). The median is displayed

as a horizontal line within the box, and the mean as a white square. Whiskers indicate the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Sample sizes: $n = 1,139$ (Y503D), $n = 24$ (K476C), $n = 7$ (wild type) runs. **d**, Translocation example for ClpB(K476C). Scale bars correspond to 200 aa and 10 s. **e**, Translocation example for wild-type ClpB with the DnaK system (DnaK, DnaJ and GrpE). Scale bars correspond to 200 aa and 5 s. **f**, **g**, Absolute ATPase rate (**f**) and ATPase substrate-stimulation (**g**) for the three ClpB variants and different substrate conditions (mean \pm s.d.). ATPase activity is higher and more strongly stimulated for Y503D, followed by K476C and wild type. The lower activities observed in the presence of denatured MBP-DM with respect to casein may reflect lower affinity and lower concentrations due to aggregation. The ATPase activity assay was repeated three times for all conditions in **f** and **g**, except for K476C, WT + MBP₂ and Y503D + casein, for which it was repeated two times.



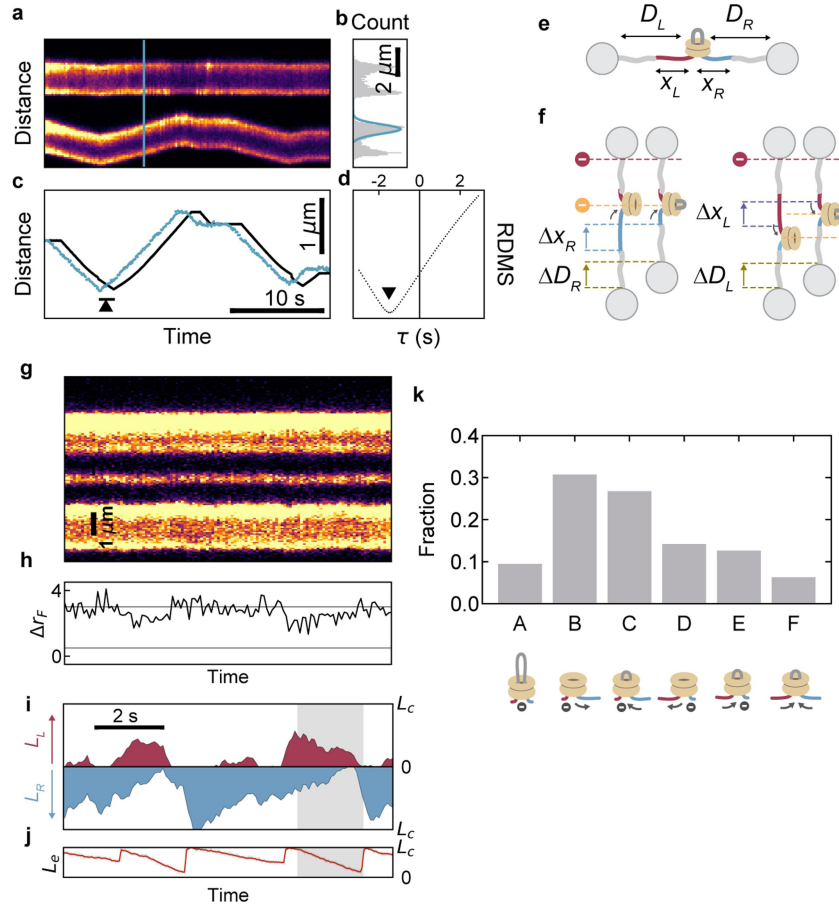
Extended Data Fig. 3 | Translocation runs for different constructs and molecules. Traces of protein extended length contractions in the presence of ClpB(Y503D) and ATP. **a**, MBP ($L_c = 360$ aa). **b**, 2MBP ($L_c = 720$ aa). **c**, 4MBP ($L_c = 1440$ aa). **d**, Speed distribution of translocation runs for the three different constructs (number of runs: $n = 213$ (MBP), $n = 287$ (2MBP), $n = 306$ (4MBP)). All show a similar range of speeds, as expected, with one main peak (at $v \approx 240$ aa s $^{-1}$) and a second peak or shoulder at twice the magnitude ($2v$). A slight change in the ratio is observed between the two peak heights, with $2v$ becoming more pronounced in the longer constructs. This difference could reflect that distances between the initial ClpB binding site and the arresting DNA handles is then larger, and hence double-arm translocation more likely (see also Extended Data Fig. 4). **e**, Translocation speed distributions from three different substrate molecules (number of runs: $n_1 = 218$ (purple), $n_2 = 102$ (orange) and $n_3 = 114$ (green)), which show no significant variability between individual substrates. **f**, Translocation speed distributions for three different translocation bursts, which show continuous run-slip-run activity, and are thus surmised to reflect the action of individual ClpB hexamers (number runs: $n_1 = 25$ (purple), $n_2 = 26$ (orange) and $n_3 = 49$ (green)). Distributions are for

ClpB(Y503D) and ATP, at approximately 8 pN. The data indicate no significant variability in the translocation activity between ClpB hexamers. The burst duration varied between 5 and 80 s, whereas the time between bursts varied between 5 and 150 s, for the 2MBP construct. **g**, Example translocation run of MBP showing the definitions of run length and run duration. Run duration is calculated as the time from the start of a run until the next back-slipping event, including the pause after translocation and before the next back-slip. Run length is calculated as the length difference between the start of a run and the next back-slipping event. **h**, **i**, Run length (**h**) and run duration (**i**) (see **g**) distribution for constructs of different lengths. Notably, the run duration distributions are similar for the constructs of different length, which suggests that the moment ClpB loses grip on one of the arms and causes the back-slip is determined by events that are intrinsic to the ClpB hexamer, and do not depend on the substrate nor the encounter with blockades (such as the DNA tether). This would make functional sense in the physiological context, as ClpB can then continue to push in an attempt to disrupt aggregated structures. By contrast, the switch between double- and single-arm translocation is directly triggered by such blockades, though without losing grip on either of the two arms.



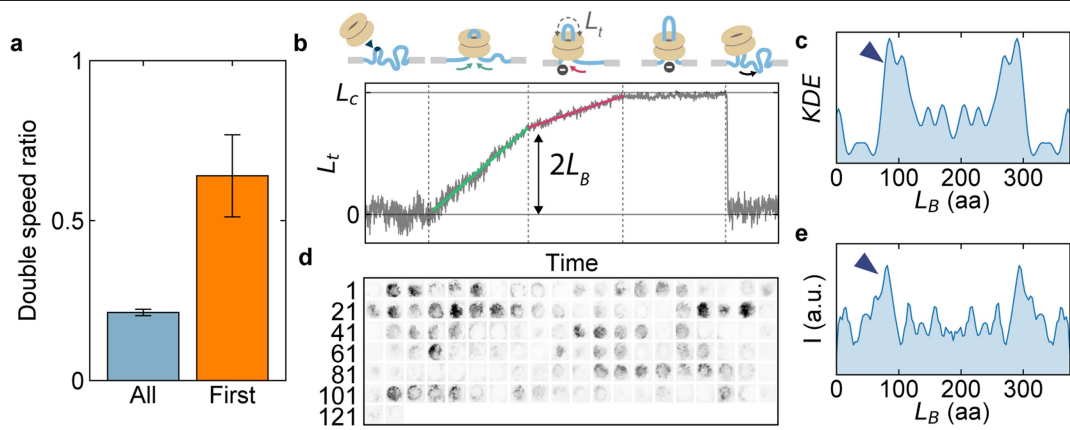
Extended Data Fig. 4 | Speed characterization of translocation runs.

a, Translocated length (L_t) during threading of 2MBP by ClpB(Y503D). Raw data (light grey, 500 Hz) is filtered using a Savitzky–Golay filter (black line). **b**, Local translocation speed calculated as the time derivative of the translocated length after Savitzky–Golay filtering. Negative slopes below -50 nm s^{-1} (horizontal line) are considered back-slipping events (orange areas, also in **a**) and help in determining isolated translocation runs. **c**, Identification of different speeds within a single translocation run. Linear fits are used to calculate the speed of the run (green and magenta lines), most times revealing two main velocities, one double that of the other. **d**, Time derivative of the filtered translocated length for a single run, with solid black lines indicating the threshold speeds to distinguish no translocation from single- and double-translocation velocities and green and magenta indicating the fitted velocity values (also shown in **c**).



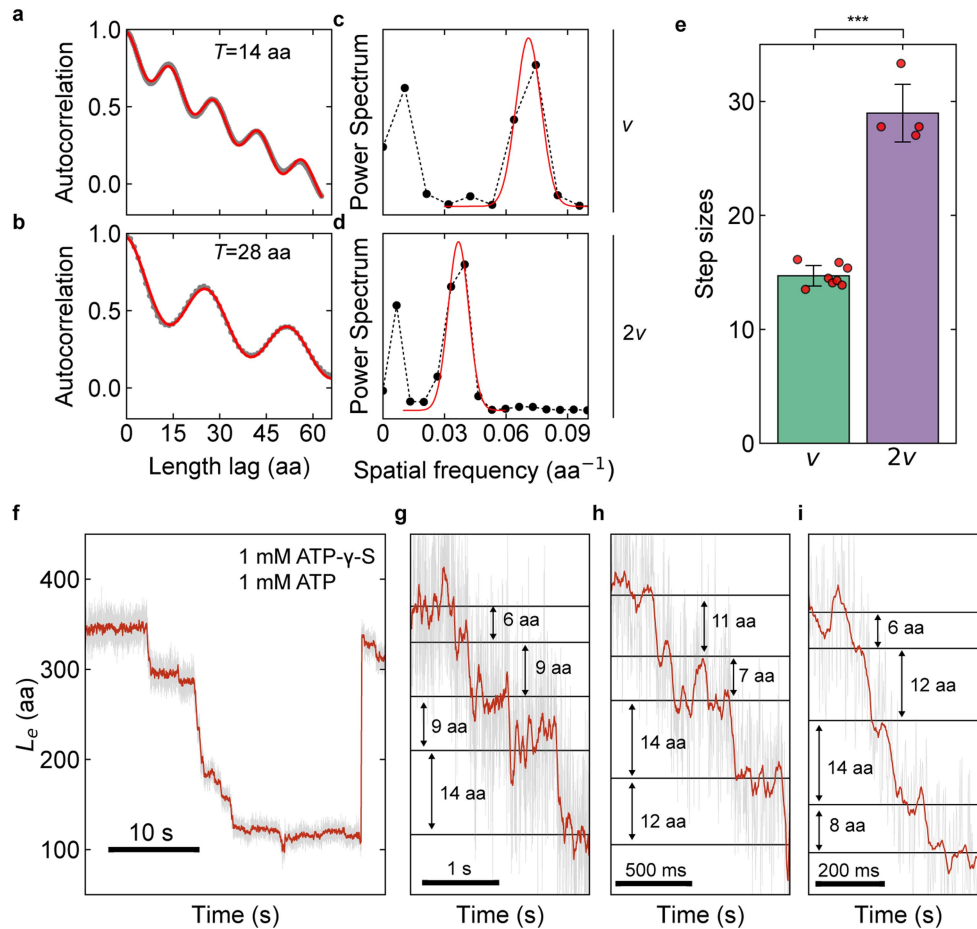
Extended Data Fig. 5 | Integrated tweezers and fluorescence particle tracking method. a–d, Synchronization of fluorescence and tweezers signals. **a**, Confocal scanning kymograph of two trapped beads. **b**, Intensity profile of a scanning line (blue in **a**), with a Gaussian fit of the edge of the moving bead (bottom) in blue. **c**, Offset between the fluorescence detection of bead movement as shown in **b** (blue dots), and high-resolution tweezers signal of trap and bead movement (black line) signals. **d**, Root mean square deviation (r.m.s.d.) between the signals for different time shifts τ . The minimum is marked with a triangle and represents the best estimation of the offset between the signals. **e, f**, Force clamp and computation of the two length components. **e**, Scheme of the lengths involved. D_L and D_R , distances between beads and ClpB; x_L and x_R , distances between protein termini and ClpB. Note that these distances are not contour lengths. **f**, Bead and ClpB position changes for left-arm (left) and right-arm (right) translocation. **g**, Kymograph underlying

data in Fig. 2i. **h**, Corresponding tracked position of ClpB. Horizontal lines indicate extreme ClpB positions. Top line, ClpB is positioned at the left-hand terminus (see **e** and **f**). Bottom line, ClpB at the right-hand terminus; no polypeptide is translocated (the complete polypeptide is thus on the *cis* side of ClpB). Deviations from the top line consistently occur at back-slip moments detected by the tweezers (**j**; see the two shorter back-slips), which shows that the left arm (red) back-slips. Some back-slips detected by the tweezers do not show a corresponding ClpB movement, which is expected because right-arm back-slips should not change the ClpB position. **i**, Corresponding lengths of left arm (red) and right arm (blue) against time, as determined from fluorescence tracking (**g, h**) and tweezers (**j**) data. **j**, Corresponding tweezers data showing the distance between termini (contour length of *cis* segments). **k**, Distribution of the different translocation and back-slipping events observed in the fluorescence experiments (number of events $n = 127$, 5 molecules).



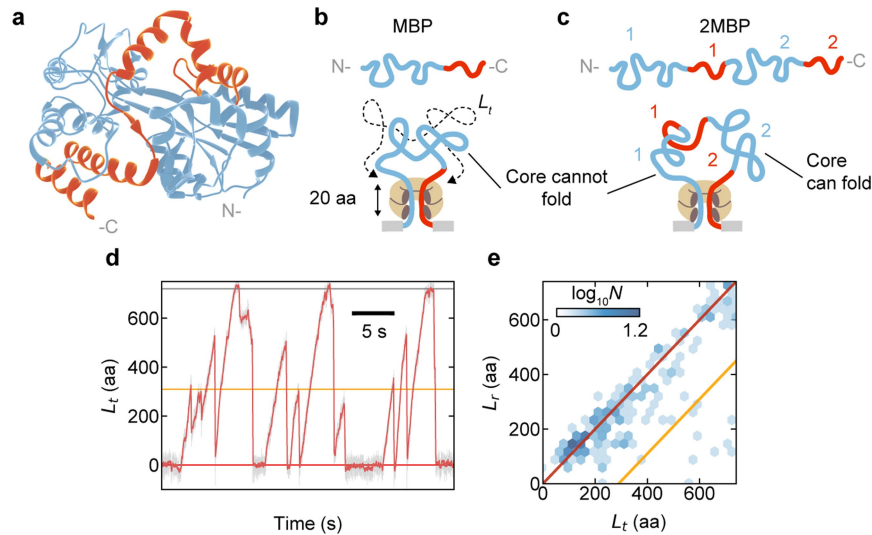
Extended Data Fig. 6 | Initial ClpB binding site estimation. **a**, Fraction of runs showing double speed when considering all runs ($n=1,704$) and first runs only ($n=30$). Data are mean \pm standard error of a binomial distribution (see Methods). **b**, Example of first translocation run. ClpB binds at a certain location on the polypeptide, starts translocating both strands yielding the double speed (green) until it encounters the closest terminus, when it switches to single strand translocation (red). At the switch, the length translocated thus equals the distance between the initial binding site and the closest terminus (L_B), but times two because ClpB also translocated the other arm. Afterwards,

the second terminus is also reached, and translocation stalls and a back-slip occurs, although this is not relevant here. **c**, Kernel density estimation (KDE) distribution of the inferred binding locations based on first runs, as described in **b** ($n=30$). The distribution is symmetric because N and C termini cannot be distinguished. **d**, Peptide library data indicating regions of MBP that are bound by ClpB(NTD). **e**, Spot intensities were quantified using a custom script in Python. For direct comparison with **c**, the spot intensity distribution was also mirrored.



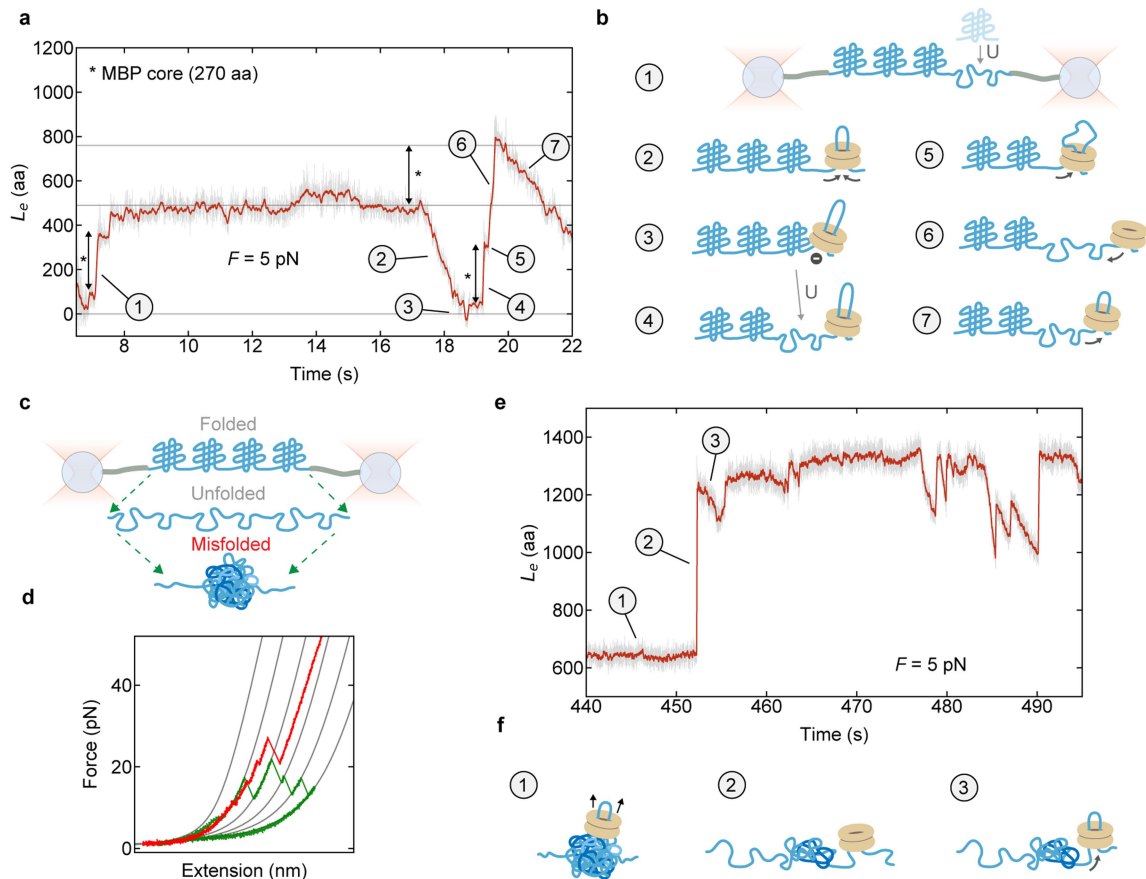
Extended Data Fig. 7 | Single translocation steps by ClpB. **a–d**, Analysis of step periodicity, related to Fig. 3. **a**, **b**, Autocorrelation of the pairwise length distribution for single-speed (**a**) and double-speed (**b**) runs from Fig. 3 (black dots). The red line is a fit, yielding period values of 14 and 28 aa, respectively. **c**, **d**, Power spectrum analysis of the pairwise length distribution for **a** (**c**) and for **b** (**d**), showing a peak that fitted to a Gaussian distribution (red) yields 0.071 and 0.037 aa^{-1} , respectively. This translates to 14 and 27 aa steps, in excellent agreement with the values obtained from the autocorrelation. **e**, The average step size is 14.6 ± 0.9 aa for single-speed translocation and 29 ± 3 aa for double-speed translocation (mean \pm s.e.m., $n_s = 8$ and $n_d = 4$, 4 molecules), and statistically different ($P = 10^{-7}$; two-sided t -test). **f**, Example run in the presence

of ATP–ATP γ S mixture (1 mM each). Longer pauses are observed during translocation because ATP γ S is hydrolysed much more slowly than ATP, and therefore can result in stalling. The prolonged stalling seen here is in line with a sequential ATP-hydrolysis along ClpB subunits. **g–i**, Notably, in these conditions, step-sizes smaller than 14 aa are now observed. These findings provide further support for the 14-aa steps being produced by the rapid consecutive action of multiple or all 6 ClpB subunits, whose individual 2-aa sub-steps would remain unresolved. After starting, a hydrolysis sequence moving along the ClpB hexamer would then arrest prematurely when encountering a ATP γ S-bound subunit, and hence yield a smaller step size.



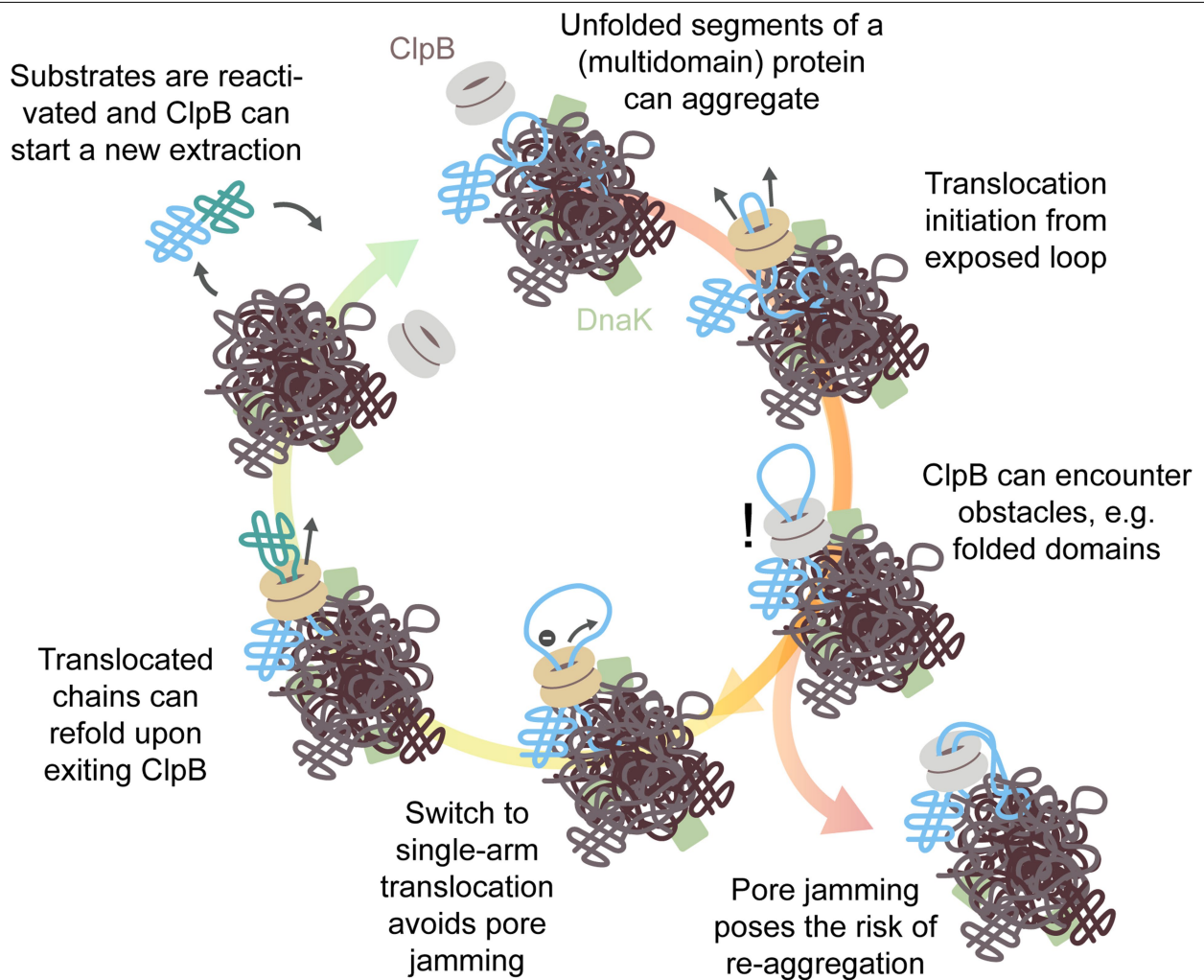
Extended Data Fig. 8 | Trans-refolding does not occur in single MBP and a mutant 2MBP construct. **a**, Structure of MBP (PDB ID: 2MV0), showing the C-terminal helices (red; around 90 residues) not required for core folding²⁸ (blue). **b**, Cartoon representation of the extended MBP chain showing the C terminal domain in red. After translocation arrest at the termini, segments at the N- and C termini (approximately 20 aa each) remain stuck inside the ClpB pore, and are thus not available for folding. Whereas the C-terminal segment (red) is not crucial for core formation, the N-terminal segment (blue) is. Thus, trans-refolding of single-MBP is not expected and indeed not observed. **c**, Cartoon representation of the extended 2MBP. The second MBP core

(blue, 2) can fold in *trans*, since it now can translocate fully. **d**, Translocation run-and-slip activity for a tandem repeat of double mutant MBP (2MBP(DM)), which is compromised in refolding. Grey line indicates 720 aa, red line corresponds to 0 aa and the orange line corresponds to 310 aa, the length of one MBP core plus the two approximately 20-aa segments inside the pore. Back-slipping arrests at the orange line, as seen for 2MBP (Fig. 4), are no longer observed. **e**, Corresponding length distribution. Upon slipping, the released length (L_r) is now typically equal to the previously translocated length (blue data follows red line, $n = 203$ runs, 6 molecules).



Extended Data Fig. 9 | Disruption of folded and aggregated structures by ClpB. **a**, Extension length (L_e) of the 4MBP construct plotted against time in the presence of ClpB(Y503D) and ATP. **b**, Cartoons of event sequence suggested in **a**. (1) One MBP core is unfolded by increasing the force, immediately followed by relaxation to 5 pN to avoid unfolding other MBP cores. Some C-terminal helices also unfolded in this process. (2) After a waiting period, ClpB binds the unfolded part and translocates it completely. (3) ClpB reaches the neighbouring folded MBP domain (and the DNA tether), and hence no longer changes L_e . (4) After a short pause, L_e increases in a discrete step of 270 aa, indicating the unfolding of one MBP core, which has precisely that length. (5) ClpB(Y503D) translocates briefly immediately afterwards, further indicating the bound ClpB, and (6) back-slipping occurs. Note that the length of the unfolded chain has increased by 270 aa, the length of one MBP core, as expected (star). (7) Translocation continues. **c**, **d**, To create a misfolded or aggregated state, the 4MBP construct was unfolded and rapidly relaxed (green trace). This sometimes produced non-native structures characterized by being

compact and highly resistant to force (red trace). The tether was then relaxed to low force. **e**, Subsequent measurement of extension length against time. **f**, Cartoons of event sequence suggested in **e**: (1) the length remains unchanged, for example, owing to waiting for ClpB binding. (2) The length increases abruptly by about 600 aa, which is more than one MBP core (270 aa), suggesting that ClpB disrupted a non-native (aggregated) structure that contained more than one MBP repeat. (3) ClpB translocation is observed immediately afterwards. This is consistent with the model, because one-step disruption of structures by ClpB (pushing) action can yield unfolded polypeptide segments directly on the *cis* side of ClpB that are then available for translocation. Note that polypeptide may also be liberated on the other side of the misfolded structure, which is not immediately available for translocation. Subsequently, further translocation and slipping behaviour is observed. Note that the structure becomes almost fully disrupted, as it nears the maximum length of 1,440 aa.



Extended Data Fig. 10 | Loop extrusion as a disaggregation principle.

Insertion and translocation of loops promotes efficient disaggregation, because aggregates may display few accessible polypeptide termini at the surface. Translocation by Hsp100s of polypeptides entangled in aggregates generates pulling forces that promote their dissociation, cooperative disruption of larger structures, and extraction. The ability of Hsp100s to switch between translocation modes is relevant to prevent pore jamming when encountering structures that resist immediate disruption. To dissolve such resistive structures and larger aggregates, many translocation actions are

probably required, involving multiple Hsp100 hexamers and other chaperones such as Hsp70, acting at different moments in time and at different locations within the aggregate. The random non-processive action of Hsp70s probably inherently requires multiple Hsp70s working together, in a manner that does not generate large pulling forces, while exploiting rapid binding and unbinding. In contrast, the processive nature of ClpB translocation enables fast, deterministic, and forced dissociation, which further limits re-aggregation and degradation when in combination with rapid refolding.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data from the Lumicks C-trap set up was acquired using the software provided by Lumicks: Tweez-O-Matic version 36.0, running under LabView 11.0. Fluorescence kymographs were acquired with Lumicks software Scanary 3.4

Data analysis

All analysis was performed using a custom made package in Python 3.5, available online upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sample sizes were chosen based on previous experience and published studies to assess reproducibility. Experiments were repeated multiple times on multiple substrate molecules, which were sufficient to obtain the described error margins.
Data exclusions	Translocation runs whose linear fits yielded r values below 0.8 were not used for translocation speed determination unless otherwise stated. This criteria was established after observing that some data contained more noise and hence were not suitable for such analysis.
Replication	All experiments were replicated multiple times, using different bead pairs and substrate molecules. All attempts at replication were successful.
Randomization	Samples were not randomized in the experiments. Randomization was not applicable as samples were allocated according to different conditions such as buffer conditions.
Blinding	Experiments were not blinded as the data acquisition and analysis were done in different conditions.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Roche Diagnostics Germany, Anti-digoxigenin bodies, Cat. No. 11333089001, polyclonal antibody from sheep
Validation	The polyclonal antibody from sheep is specific to digoxigenin and digoxin and shows no cross-reactivity with other steroids, such as human estrogens and androgens .

The molecular basis for sugar import in malaria parasites

<https://doi.org/10.1038/s41586-020-1963-z>

Received: 17 June 2019

Accepted: 3 January 2020

Published online: 29 January 2020

Abdul Aziz Qureshi^{1,3}, Albert Suades^{1,3}, Rei Matsuoka¹, Joseph Brock¹, Sarah E. McComas^{1,2}, Emmanuel Nji¹, Laura Orellana¹, Magnus Claesson¹, Lucie Delemotte² & David Drew^{1*}

Elucidating the mechanism of sugar import requires a molecular understanding of how transporters couple sugar binding and gating events. Whereas mammalian glucose transporters (GLUTs) are specialists¹, the hexose transporter from the malaria parasite *Plasmodium falciparum* PfHT1^{2,3} has acquired the ability to transport both glucose and fructose sugars as efficiently as the dedicated glucose (GLUT3) and fructose (GLUT5) transporters. Here, to establish the molecular basis of sugar promiscuity in malaria parasites, we determined the crystal structure of PfHT1 in complex with D-glucose at a resolution of 3.6 Å. We found that the sugar-binding site in PfHT1 is very similar to those of the distantly related GLUT3 and GLUT5 structures^{4,5}. Nevertheless, engineered PfHT1 mutations made to match GLUT sugar-binding sites did not shift sugar preferences. The extracellular substrate-gating helix TM7b in PfHT1 was positioned in a fully occluded conformation, providing a unique glimpse into how sugar binding and gating are coupled. We determined that polar contacts between TM7b and TM1 (located about 15 Å from D-glucose) are just as critical for transport as the residues that directly coordinate D-glucose, which demonstrates a strong allosteric coupling between sugar binding and gating. We conclude that PfHT1 has achieved substrate promiscuity not by modifying its sugar-binding site, but instead by evolving substrate-gating dynamics.

P. falciparum relies on a continuous supply of host-derived glucose during the clinically important asexual stages of growth and replication within erythrocytes⁶. As a consequence, glucose consumption is increased about 100-fold in erythrocytes that contain the parasite⁷. This metabolism is further dependent upon the import of glucose across the cell membrane of the parasite by the hexose transporter PfHT1^{2,3} (Fig. 1a). Owing to its essential role in glucose metabolism, PfHT1 is a well-recognized target for antimalarial drugs^{8–11}. PfHT1 belongs to the major facilitator superfamily (MFS), the members of which share a fold that consists of two symmetrical six transmembrane (TM) bundles of helices^{12,13}—as was first clearly observed in the structure of lactose permease (LacY)¹⁴. However, PfHT1 clusters with a separate MFS class than that of LacY; it belongs to the subfamily of sugar porters^{15,16}, which includes the medically relevant GLUT transporters¹. In contrast to the GLUT transporters (which show poor catalytic activity for diverse sugars¹), PfHT1 shows a broader substrate specificity^{2,17}. In particular, PfHT1 has acquired the ability to transport both D-glucose and D-fructose with kinetics (K_M) similar to those of the dedicated high-affinity D-glucose (GLUT3) and D-fructose (GLUT5) transporters, respectively^{1–3}. Structures of the related sugar porters GLUT1¹⁸, GLUT3⁴ and GLUT5⁵, as well as the *Escherichia coli* xylose transporter XylE^{19–21}, have previously been determined^{22,23}. Because PfHT1 shares only low sequence identity with these transporters (Extended Data Fig. 1a, b), it has been unclear whether sugar recognition would be similar. Here we aimed to determine the structure

of PfHT1 to establish the molecular ‘rules’ that govern its substrate specificity and inhibition.

Purified PfHT1 was reconstituted into liposomes and showed robust uptake of radiolabelled D-glucose, D-mannose, D-galactose, D-fructose and D-xylose, consistent with in vivo analysis³ (Fig. 1b, Extended Data Fig. 2a–e). PfHT1 can also transport D-glucosamine (Extended Data Fig. 2f), as has previously been observed for GLUT2 and the galactose transporter GalP^{24,25}. PfHT1 kinetics for D-glucose (K_M of 0.80 mM) and D-fructose (K_M of 9.6 mM) in proteoliposomes was comparable to in vivo estimates for PfHT1, GLUT3 and GLUT5^{1,3,26,27} (Extended Data Fig. 3a, d). The turnover rates (k_{cat}) for D-glucose (19 s^{−1}) and D-fructose (30 s^{−1}) were further comparable to in vitro estimates for those of GLUT3 (13 s^{−1})²⁸ and GLUT5 (43 s^{−1}), respectively (Extended Data Table 1). We co-crystallized PfHT1 with D-glucose using the vapour-diffusion method, and determined the structure by molecular replacement at a resolution of around 3.6 Å (Fig. 1c, Extended Data Table 2). PfHT1 crystallized as a dimer, with four molecules in the asymmetric unit (Extended Data Fig. 4a). The PfHT1 structure is highly similar to the outward-occluded structure of human GLUT3⁴ (Extended Data Fig. 4b). The extracellular half-helix TM7b in PfHT1 was found to be more occluded than in human GLUT3, and matched the position of TM7b in the inward-open conformation of GLUT1 and GLUT5 (Fig. 1d, Extended Data Fig. 4d). Nonetheless, PfHT1 was not in an inward-facing state as access to the inside was closed. We conclude PfHT1 has been captured in a previously unobserved, fully occluded conformation (Fig. 1c, e).

¹Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. ²Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. ³These authors contributed equally: Abdul Aziz Qureshi, Albert Suades. *e-mail: ddrew@dbb.su.se

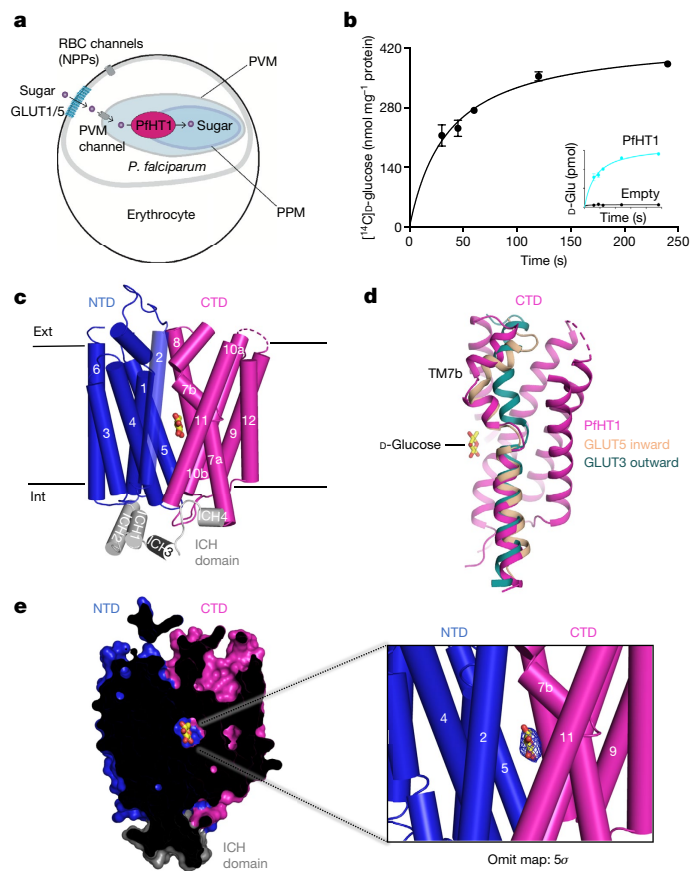


Fig. 1 | PfHT1 in D-glucose-bound occluded conformation. **a**, *P. falciparum* infected erythrocyte illustrating the localization of PfHT1 within the parasitic plasma membrane and the GLUT-dependent uptake of glucose and fructose. GLUT1/5, GLUT1 or GLUT5; NPP, new permeability pathway; RBC, red blood cell; PVM, parasitophorous vacuole membrane; PPM, parasite plasma membrane. **b**, Time-dependent uptake of [¹⁴C]D-glucose (black circles) by PfHT1 proteoliposomes. Inset, PfHT1 [¹⁴C]D-glucose (cyan trace) and non-specific uptake in empty liposomes (black trace). Error bars represent mean ± s.e.m. of *n* = 3 biologically independent experiments. D-Glu, D-glucose. **c**, Cartoon representation of the structure of the PfHT1 D-glucose-bound complex in the occluded conformation, showing the N-terminal six-transmembrane (6TM) domain (NTD) (blue), the C-terminal 6TM domain (CTD) (magenta), the intrahelical domain (ICH) (grey) and D-glucose (stick representation). PfHT1 has a large intracellular salt-bridge network that stabilizes the occluded conformation, as seen in structures of related sugar porters in outward-facing conformations^{22,23} (Extended Data Fig. 4c). The structures of sugar porters in the outward-facing conformation are further stabilized by interactions between the intracellular helices ICH1, ICH2, ICH3 and ICH4, which collectively interact with ICH5 to latch the NTD and CTD together⁴. ICH5 has previously been observed only in the outward-facing conformation^{4,5}. In the occluded PfHT1 structure, we were unable to model ICH5 (Extended Data Fig. 4e), which implies that PfHT1 is primed for transition into the inward-facing conformation. Ext, exterior; int, interior. **d**, Ribbon representation of the CTD domain of PfHT1 (magenta), superimposed with TM7 and TM10 gating helices of outward-facing GLUT3 (RCSB Protein Data Bank code (PDB) 4ZWC, shown in teal) and inward-facing GLUT5 (PDB 4YB9, shown in light orange). **e**, Surface representation of the PfHT1 structure in the occluded conformation with D-glucose shown as sticks (left), and the corresponding polder omit map (blue mesh at 5σ) shown for the D-glucose surrounded by the NTD and CTD helices (right), coloured as in c.

We observed considerable non-protein electron density in the C-terminal bundle of PfHT1 that corresponded exactly to the sugar-binding site for D-glucose in human GLUT3⁴ (Figs. 1e, 2a, b, Extended Data Fig. 4e). Almost all D-glucose hydrogen-bonding residues in human

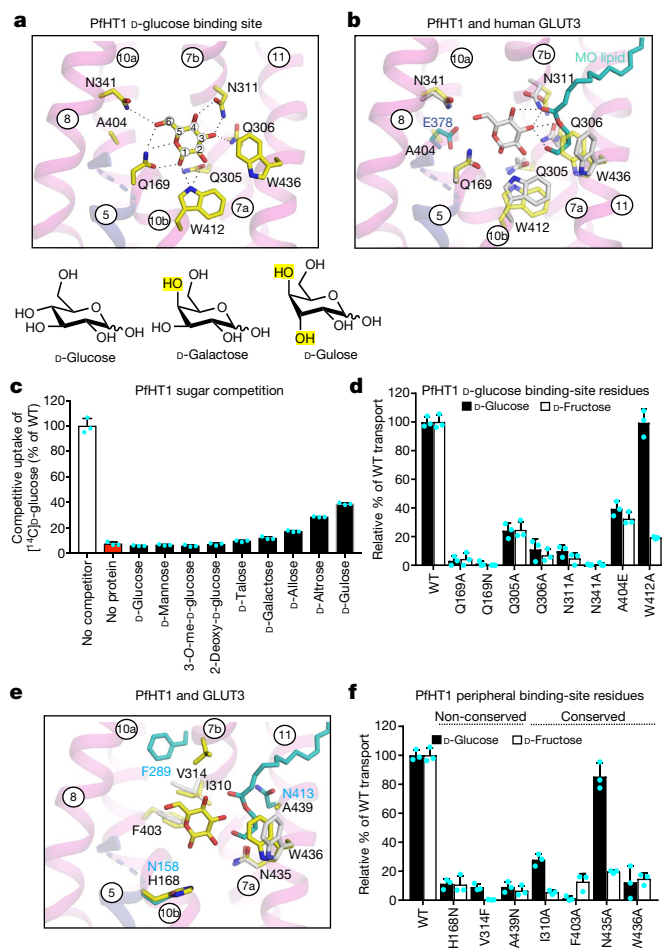


Fig. 2 | Molecular recognition of D-glucose by PfHT1. **a**, Cartoon representation of the PfHT1 sugar-binding site with D-glucose (yellow sticks) and the interacting residues labelled; residues Q169, Q305, Q306, N311 and N341 were determined to be essential for sugar transport, as shown in d. Putative hydrogen bonds are indicated with dotted lines. **b**, Sugar-binding site comparison between PfHT1 (yellow sticks) and human GLUT3 (PDB 4ZW9) (grey, conserved; cyan, non-conserved). The crystallization lipid monoolein (MO) interacting with TM7b in human GLUT3 is shown in cyan, and the PfHT1 A404 residue that corresponds to E378 in human GLUT3 is shown in dark blue. Putative hydrogen bonds are indicated with dotted lines. **c**, The competitive uptake of [¹⁴C]D-glucose by PfHT1 proteoliposomes in the absence (white bar) and presence of non-labelled D-glucose epimers and homologues (black bars); note, competitive uptake cannot distinguish between transported and non-transported sugars and non-specific uptake in empty liposomes (red bar). Data are mean ± s.e.m. of *n* = 3 biologically independent experiments. **d**, Transport activity of PfHT1 mutants for residues in D-glucose-binding site, for [¹⁴C] D-glucose (black bars) and [¹⁴C] D-fructose (white bars). Data are mean ± s.e.m. of *n* = 3 biologically independent experiments. **e**, Comparison of peripheral D-glucose-binding site between PfHT1 (yellow sticks) and human GLUT3 (grey, conserved; cyan, non-conserved), and monoolein lipid interacting with human GLUT3 in cyan. **f**, Transport activity of PfHT1 mutants for residues in the peripheral binding site (in which the residue is substituted with alanine or the equivalent residue in human GLUT3) for [¹⁴C] D-glucose (black bars) and [¹⁴C] D-fructose (white bars). Data are mean ± s.e.m. of *n* = 3 biologically independent experiments.

GLUT3 were conserved and similarly positioned in PfHT1 (Fig. 2a, b, Extended Data Fig. 1c). D-Glucose in PfHT1 was therefore modelled with the observed orientation in human GLUT3 and refined with appropriate stereochemical restraints (Methods). It was, nonetheless, important to validate the binding pose of D-glucose, especially since a crystallization lipid interacted with D-glucose in human GLUT3 (Fig. 2b). Forthwith,

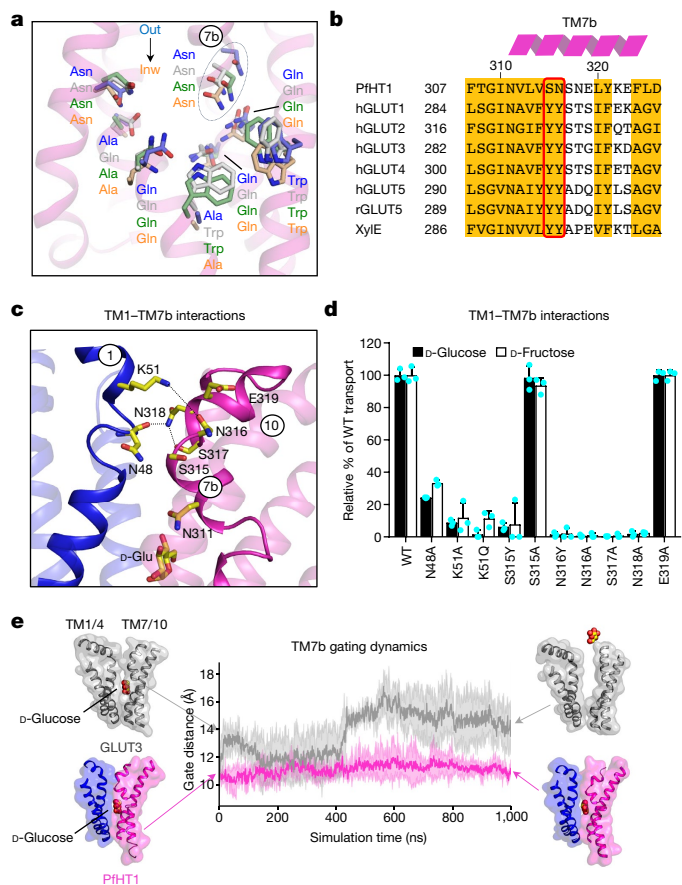


Fig. 3 | Gating helix in PfHT1 enables substrate promiscuity. **a**, Cartoon representation of the sugar-binding site in occluded PfHT1 (green sticks) superimposed with outward-open GLUT5 (PDB 4YB9, orange sticks), outward-occluded GLUT3 (PDB 4ZW9, grey sticks) and inward-open GLUT1 (PDB 4YBQ, blue sticks). Asn311 (dotted ellipsoid) is the only residue that clearly repositions during the entire transport cycle. **b**, TM7b sequence alignment between human (h) GLUT1, GLUT2, GLUT3, GLUT4 and GLUT5, rat (r) GLUT5, *E. coli* XyleE and PfHT1. The red box highlights that the highly conserved occlusion-forming tyrosine residues²³ are replaced by serine and asparagine in PfHT1. The yellow shading highlights conserved residues. **c**, Cartoon representation of PfHT1 extracellular gating interactions between TM7b (magenta) and TM1 (blue). Potential hydrogen-bond interactions are indicated by dotted lines and prominent residue side chains are labelled. **d**, Transport activity for TM1-TM7b interacting-residue mutants for [¹⁴C]D-glucose (black bars) and [¹⁴C]D-fructose (white bars). Residues K51, N311, N316, S317 and N318 were determined to be essential for sugar transport; their relative positions are shown in **c**. Data are mean \pm s.e.m. of $n = 3$ biologically independent experiments. **e**, Gating interactions for the D-glucose-bound outward-occluded structure of human GLUT3 (grey) (top) and the D-glucose-bound occluded structure of PfHT1 (bottom). Blue, NTD; magenta, CTD; shaded lines represent the distribution of gating distances from $n = 3$ independent 1- μ s molecular dynamics simulations (Methods) and non-shaded lines represent their respective mean distance. Representative side views for human GLUT3 and PfHT1 at the start and end of the simulation time are also shown.

[¹⁴C]D-glucose uptake by PfHT1 proteoliposomes was performed in the presence of unlabelled epimers of D-glucose and homologous sugars (Fig. 2c, Extended Data Fig. 5a, b). Most revealingly, D-glucose—which differs from D-galactose in only the C3-hydroxyl orientation—displayed fivefold-poorer competition for [¹⁴C]D-glucose uptake than D-galactose. Consistently, D-allose (the C3 epimer of D-glucose) was clearly weaker at competing for [¹⁴C]D-glucose uptake than either the C4 epimer D-galactose or the C2,C4 epimer D-talose. Moreover, D-mannose and

2-deoxyglucose, which differ from D-glucose in their C2 positions, were just as competitive for [¹⁴C]D-glucose uptake as D-glucose. Overall, the C3-hydroxyl group orientation was determined to be the most critical for D-glucose recognition, followed by the C4-hydroxy group orientation.

In the PfHT1 structure, the C3- and C4-hydroxyl groups hydrogen bond to Asn311 and Gln306, whereas the C1- and C2-hydroxyl groups hydrogen bond to Gln169, Gln305 and Trp412 (Fig. 2a). Consistently, whereas alanine substitutions of Trp412 and Gln305 residues retained 100% and 22% of wild-type activity (respectively), Gln306 and Asn311 alanine mutations were less than 10% active (Fig. 2d, Extended Data Fig. 6a). In the observed orientation, the C1- and C2-hydroxyl groups face the interior and the C3- and C4-hydroxyl groups face the exterior, which is consistent with the binding pose that is biochemically predicted for GLUT1²⁹. Because *E. coli* XyleE also binds D-glucose in a manner similar to that of PfHT1 and GLUT3¹⁹, the D-glucose binding mode appears to be evolutionarily conserved (Extended Data Fig. 6d). Notably, the strictly conserved Gln169 (in TM5) is the only N-terminal-bundle residue that coordinates D-glucose and is—as expected—also critical for transport (Fig. 2d).

To investigate PfHT1 promiscuity, we focused on the ability of PfHT1 mutants to transport D-glucose versus D-fructose, which are the most physiologically relevant sugars and have sevenfold differences in specificity constants (k_{cat}/K_M) (Extended Data Fig. 3f, g). We substituted Ala404 with glutamate, as this residue was the only obvious difference between the human GLUT3 and PfHT1 sugar-binding sites (Fig. 2b). However, the Ala404Glu mutant retained D-glucose and D-fructose transport (Fig. 2d, Extended Data Fig. 6a, e, f). We next investigated whether residues peripheral to the main sugar-binding site might influence substrate selectivity (Fig. 2e, f). Consequently, we generated His168Asn, Val314Phe and Ala439Asn single mutants of PfHT1 to mimic human GLUT3; in these mutants, transport of both D-glucose and D-fructose were again similarly impaired (Fig. 2f, Extended Data Fig. 6b). As a comparison, we also assayed alanine substitutions of the conserved residues Ile310, Phe403, Asn435 and Trp436; these substitutions also led to impaired transport of both sugars—with the exception of Asn435Ala, which selectively abolished D-fructose transport (Fig. 2f, Extended Data Fig. 6b).

As we were unable to rationalize the sugar preferences of PfHT1 on the basis of human GLUT3, we extended our comparison to rat GLUT5, which revealed that the Trp412 in PfHT1 is replaced by alanine in GLUT5 (Fig. 2b, Extended Data Fig. 6g). However, the Trp412Ala mutant retained D-glucose transport but had severely reduced D-fructose transport (Fig. 2d, Extended Data Fig. 6h). Taken together, our experiments suggest that D-fructose transport in PfHT1 requires almost the same set of sugar-binding residues as does D-glucose transport, but that the former shows greater sensitivity to mutagenesis—probably because it interacts with lower affinity^{3,17}. Indeed, the Asn435 and Trp412 alanine mutations that selectively affect D-fructose transport (Fig. 2d, f) nevertheless have reduced D-glucose turnover and are not found in transporters that are specific for D-fructose (Extended Data Table 1, Extended Data Figs. 1c, 6a, b). It was thus unclear how PfHT1 robustly transports different sugars.

Despite the high levels of structural similarity, two antimalarial compounds C3361 and MMV009085 have previously been discovered to have 19- to 250-fold higher selectivity for PfHT1 over human GLUT1, GLUT5, and over GLUT1–4, respectively^{8,11,30}. The compound C3361 is a D-glucose derivative with an undec-10-en- addition at the C3-hydroxyl position; aliphatic chain additions to the C3 hydroxyl showed the strongest inhibition (followed by additions at the C4 hydroxyl), whereas additions to C1, C5 or C6 positions showed no inhibition⁸. In the occluded PfHT1 structure, there is a narrow hydrophobic vestibule that would be accessible only from the C3- and C4-hydroxyl positions (Extended Data Fig. 7a). In the glucose-bound human GLUT3 structure, a crystallization lipid consistently occupies this site (Extended Data

Fig. 7a). Although the binding mode of MMV009085 was previously unknown, we conclude that it also binds in the sugar-binding pocket because—similar to cytochalasin B—PfHT1 inhibition was found to be dependent on Trp412 (Extended Data Fig. 7b). The malarial box inhibitor MMV009085 is a symmetric tetracyclic compound with two butanol moieties (Extended Data Fig. 7c). Consistent with inhibition requiring interaction with the hydrophobic-side vestibule, a compound synthesized without the butanol moieties was unable to inhibit PfHT1 (Extended Data Fig. 7c). Thus, the occluded structure highlights how an off-site vestibule could be targeted to improve selective inhibition of PfHT1.

By combining the occluded PfHT1 structure with previous sugar-porter structures, we can reconstruct what is arguably the most complete MFS transporter cycle known to date (Supplementary Video 1). During the rocker-switch alternating-access mechanism²², the N- and C-terminal bundles clearly rearrange around the centrally located substrate-binding site. Global rearrangements are further coupled with local rearrangements of the TM7b and TM10a half-helices that gate access to the sugar-binding site from the outside and inside, respectively^{4,5,22}. Side-chain positioning of almost all sugar-binding residues are virtually unchanged during the entire transport cycle (Fig. 3a), which implies sugar translocation must be primarily driven by conformational selection. The residue Asn311 in the extracellular substrate-gating helix TM7b is the only residue that moves substantially during the transport cycle. Specifically, in the transition to a sugar-bound occluded state, Asn311 moves inward to form hydrogen bonds with the critically important C3- and C4-hydroxyl groups (Fig. 3a, Extended Data Fig. 8a). In human GLUT3, two highly conserved TM7b tyrosine residues move in concert with Asn311 to occlude sugar exit⁴ (Extended Data Fig. 8a). Thus, the strictly conserved Asn311 is probably a generic interaction site that couples sugar binding to TM7b gating.

The occlusion-forming tyrosine residues that are strictly conserved in the GLUT proteins are replaced in PfHT1 by the polar residues Ser315 and Asn316, and the substitution of either with tyrosine abolished transport (Fig. 3b, d, Extended Data Fig. 6c). Asn316 extended towards TM1 and seemed to form polar interactions with Lys51 (Fig. 3c). The mutation of Lys51, which is located about 15 Å from D-glucose, to alanine or glutamine also rendered PfHT1 non-functional (Fig. 3d, Extended Data Fig. 6c). TM7b and TM1 interactions were further observed between the backbone oxygen of Asn48 in TM1 and Asn318 in TM7b, which formed hydrogen bonds with Ser317 (Fig. 3c, Extended Data Fig. 6c). Ser317Ala and Asn318Ala (which affect TM7b) mutants also rendered PfHT1 non-functional, whereas the Asn48Ala mutant had severely reduced activity (Fig. 3d, Extended Data Fig. 6c). By contrast, alanine mutations of Ser315 and Glu319 residues (which point away from the TM1–TM7b interface) retained robust transport activity (Fig. 3c, d). The functional Ser315Ala and Glu319Ala mutants nevertheless show a reduction in turnover, mostly for D-glucose (Extended Data Table 1). To probe TM1 and TM7b interactions further, we compared 1-μs molecular dynamics simulations of human GLUT3 and PfHT1 structures. In the presence or absence of D-glucose, TM7b in human GLUT3 was found to be very mobile and it consistently moved far enough apart from TM1 to enable the release of D-glucose (Fig. 3e, Extended Data Fig. 8b). By contrast, TM7b in PfHT1 mostly retained an occluded conformation (Fig. 3e, Extended Data Fig. 8c) but was somewhat more mobile in the absence of D-glucose. For the majority of the simulation time, TM1 and TM7b contacts were maintained between the Lys51 and Asn316 residues (Extended Data Fig. 8d). Taken together, our findings show that TM1 and TM7b gating interactions and dynamics appear to be of equal importance to sugar transport kinetics as the residues in the sugar-binding site.

The formation of the occluded state is an important intermediate for understanding substrate coupling. A statistical comparison of the PfHT1 structure supports its designation as an occluded conformation that links previously determined sugar-porter states (Fig. 4, Extended Data Fig. 9a–c). For a sugar to be a substrate, it not only has to bind but

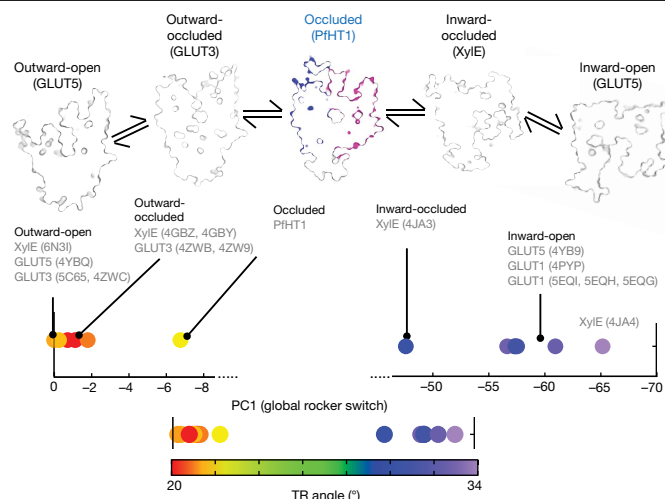


Fig. 4 | The conformational-selection-driven rocker-switch mechanism for facilitative sugar transport. The fully occluded conformation of PfHT1 is the last remaining state to be observed within the rocker-switch alternating-access mechanism of MFS transporters that belong to the sugar-porter subfamily. The observed structural states are shown as surface transversal cross-sections and clockwise from the top left: outward-open rat GLUT5 (PDB 4YBQ), outward-occluded human GLUT3 (PDB 4ZW9), fully occluded PfHT1, inward-occluded Xyle (PDB 4JA3) and inward-open bovine GLUT5 (PDB 4YB9). In the forward or reverse direction, the attainment of the occluded intermediate represented by PfHT1 is required. The principal component (PC) analysis from the conserved MFS ensemble core ($n = 17$ structures from 16 PDB codes; Methods) is shown below the structures; this analysis yields a major PC1 component (65% of the total structural variance) that tracks the rocker-switch global motion. Projections are coloured according to the angle between tandem repeats (TR) (Methods).

must also induce formation of the occluded state, which is a prerequisite for alternating access. In the transition from the outward-occluded to the fully occluded conformation, TM7b breaks and extends closer to TM1, adopting the position seen in inward-facing structures (Extended Data Figs. 4d, 8a). The fact that the occluded state can be observed by crystallography implies this state is likely to be more stable in PfHT1 than it is in GLUT proteins, consistent with the comparative molecular dynamics simulations and the additional polar interactions observed between the TM7b and TM1 helices. Rather than modifying the chemistry of the sugar-binding pocket, we conclude PfHT1 has evolved TM7b substrate-gating dynamics so that it can transition into the occluded state more easily. In this way, the PfHT1 is a more robust and promiscuous sugar transporter than the GLUT transporters, as it is less sensitive to a specific type of sugar being bound. Certainly, substrate promiscuity would be an advantage to *P. falciparum*, which is able to use D-glucose or D-fructose as a sole source of energy^{2,3,8}.

Although TM7b substrate-gating interactions and dynamics might be exaggerated in PfHT1, we think they are also of functional importance to GLUT proteins. Indeed, the QLS motif—which (prior to structural information) was thought to confer D-fructose specificity by acting as a selectivity filter—is not located in the main sugar-binding pocket, but is instead juxtaposed to the TM7a and TM7b breakpoint^{4,31}. Likewise, an isoleucine-to-valine mutation in GLUT7 that abolishes D-fructose transport (while leaving D-glucose transport unaffected) is not located in the sugar-binding pocket, but is instead between TM7b and TM10 half-helices³². The importance of fine-tuned sugar-binding and gating would further explain why Xyle binds D-glucose in a manner similar to that of PfHT1 and human GLUT3 (Extended Data Fig. 6d) but is incapable of transporting the sugar. To conclude, PfHT1 highlights that substrate-gating dynamics is probably a greater determinant for evolving sugar

translocation than previously thought, and should be considered more closely in establishing transport mechanisms in general.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1963-z>.

- Mueckler, M. & Thorens, B. The SLC2 (GLUT) family of membrane transporters. *Mol. Aspects Med.* **34**, 121–138 (2013).
- Woodrow, C. J., Penny, J. I. & Krishna, S. Intraerythrocytic *Plasmodium falciparum* expresses a high affinity facilitative hexose transporter. *J. Biol. Chem.* **274**, 7272–7277 (1999).
- Woodrow, C. J., Burchmore, R. J. & Krishna, S. Hexose permeation pathways in *Plasmodium falciparum*-infected erythrocytes. *Proc. Natl Acad. Sci. USA* **97**, 9931–9936 (2000).
- Deng, D. et al. Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **526**, 391–396 (2015).
- Nomura, N. et al. Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* **526**, 397–401 (2015).
- Kirk, K., Horner, H. A. & Kirk, J. Glucose uptake in *Plasmodium falciparum*-infected erythrocytes is an equilibrative not an active process. *Mol. Biochem. Parasitol.* **82**, 195–205 (1996).
- Roth, E. Jr. *Plasmodium falciparum* carbohydrate metabolism: a connection between host cell and parasite. *Blood Cells* **16**, 453–460, 461–466 (1990).
- Joet, T., Eckstein-Ludwig, U., Morin, C. & Krishna, S. Validation of the hexose transporter of *Plasmodium falciparum* as a novel drug target. *Proc. Natl Acad. Sci. USA* **100**, 7476–7479 (2003).
- Dean, P., Major, P., Nakjang, S., Hirt, R. P. & Embley, T. M. Transport proteins of parasitic protists and their role in nutrient salvage. *Front Plant Sci* **5**, 153 (2014).
- Ortiz, D. et al. Identification of selective inhibitors of the *Plasmodium falciparum* hexose transporter PfHT by screening focused libraries of anti-malarial compounds. *PLoS ONE* **10**, e0123598 (2015).
- Krishna, S. et al. Transport processes in *Plasmodium falciparum*-infected erythrocytes: potential as new drug targets. *Int. J. Parasitol.* **32**, 1567–1573 (2002).
- Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **38**, 151–159 (2013).
- Madej, M. G., Sun, L., Yan, N. & Kaback, H. R. Functional architecture of MFS D-glucose transporters. *Proc. Natl Acad. Sci. USA* **111**, E719–E727 (2014).
- Abramson, J. et al. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
- Maiden, M. C., Davis, E. O., Baldwin, S. A., Moore, D. C. & Henderson, P. J. Mammalian and bacterial sugar transport proteins are homologous. *Nature* **325**, 641–643 (1987).
- Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
- Blume, M. et al. A constitutive pan-hexose permease for the *Plasmodium* life cycle and transgenic models for screening of antimalarial sugar analogs. *FASEB J.* **25**, 1218–1229 (2011).
- Deng, D. et al. Crystal structure of the human glucose transporter GLUT1. *Nature* **510**, 121–125 (2014).
- Sun, L. et al. Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature* **490**, 361–366 (2012).
- Quistgaard, E. M., Löw, C., Moberg, P., Trésaugues, L. & Nordlund, P. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nat. Struct. Mol. Biol.* **20**, 766–768 (2013).
- Wisedchaisri, G., Park, M. S., Iadanza, M. G., Zheng, H. & Gonen, T. Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nat. Commun.* **5**, 4521 (2014).
- Drew, D. & Boudker, O. Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem.* **85**, 543–572 (2016).
- Yan, N. A glimpse of membrane transport through structures—advances in the structural biology of the GLUT glucose transporters. *J. Mol. Biol.* **429**, 2710–2725 (2017).
- Uldry, M., Ibberson, M., Hosokawa, M. & Thorens, B. GLUT2 is a high affinity glucosamine transporter. *FEBS Lett.* **524**, 199–203 (2002).
- Majid, H. et al. Screening of candidate substrates and coupling ions of transporters by thermostability shift assays. *eLife* **7**, e38821 (2018).
- Colville, C. A., Seatter, M. J., Jess, T. J., Gould, G. W. & Thomas, H. M. Kinetic analysis of the liver-type (GLUT2) and brain-type (GLUT3) glucose transporters in *Xenopus* oocytes: substrate specificities and effects of transport inhibitors. *Biochem. J.* **290**, 701–706 (1993).
- Burant, C. F., Takeda, J., Brot-Laroche, E., Bell, G. I. & Davidson, N. O. Fructose transporter in human spermatozoa and small intestine is GLUT5. *J. Biol. Chem.* **267**, 14523–14526 (1992).
- Hresko, R. C., Kraft, T. E., Quigley, A., Carpenter, E. P. & Hruz, P. W. Mammalian glucose transporter activity is dependent upon anionic and conical phospholipids. *J. Biol. Chem.* **291**, 17271–17282 (2016).
- Holman, G. D. Chemical biology probes of mammalian GLUT structure and function. *Biochem. J.* **475**, 3511–3534 (2018).
- Kraft, T. E. et al. A novel fluorescence resonance energy transfer-based screen in high-throughput format to identify inhibitors of malarial and human glucose transporters. *Antimicrob. Agents Chemother.* **60**, 7407–7414 (2016).
- Seatter, M. J., De la Rue, S. A., Porter, L. M. & Gould, G. W. QLS motif in transmembrane helix VII of the glucose transporter family interacts with the C-1 position of D-glucose and is involved in substrate selection at the exofacial binding site. *Biochemistry* **37**, 1322–1326 (1998).
- Manolescu, A., Salas-Burgos, A. M., Fischbarg, J. & Cheeseman, C. I. Identification of a hydrophobic residue as a key determinant of fructose transport by the facilitative hexose transporter SLC2A7 (GLUT7). *J. Biol. Chem.* **280**, 42978–42983 (2005).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Construct design and cloning

PfHT1 was cloned into the GAL-inducible vector pDDGFP2. The resulting construct consisted of: residues 1–5 from rat GLUT5 to facilitate recombinant expression, PfHT1 residues 20–504 (out of 504) (UniProt accession number: O97467), followed by a tobacco etch virus (TEV) cleavage site and a C-terminal GFP–His₈ tag. The vector was transformed into the *Saccharomyces cerevisiae* strain FGY217 (MAT α , ura3–52, lys2 Δ 201 and pep4 Δ)³³ as previously described³⁴. The resulting translated sequence following TEV digestion, with the non-PfHT1 residues from GLUT5 and residues of the TEV cleavage site underlined, is: MEKEDSGFFSTSFKYVLSACIASFIFGYQVSVLNTIKNFIVVEFEWCKG EKDRLNCSNNTIQSSFLASVFIGAVLGCGSGYLVQFGRRLSLIIYNFFFLV SILTSITHHFHTILFARLLSGFGIGLVTSVPMYISEMTHKDKKGAYGVMHQL FITFGIFVAVMLGLAMGEGPKADSTEPLTSFAKLWRLMFLFPSVISLIGIL ALVVFFKEETPYFLFEKGRIEESKNILKKIYETDNVDEPLNAIKEAVEQNESA KKNLSLSLLSALKIPSYRYVILGCLLSGLQQTGINVLVSNNSNELYKEFLDSH LITLSVVMATAVNFMTFAPIYIVEKLGRKTLLWGCVGLVAYLPTAIANEI NRNSNFVKILSIVATFVMIISFAVSYGVPVLWIYHEMFPEIKDSASLASLV NWVCAIIVFPDIIIKKSPSILFIVFSVMSILTFFFIFFIKETKGGEIGTSPYIT MEERQKHMTKSVVENLYFQ.

Large-scale production and purification

For large-scale production, 24 l of *S. cerevisiae* FGY217 cells were grown in –URA medium containing 0.1% (v/v) glucose at 30 °C in 2-l shaking flasks. Protein production was induced at an optical density at 600 nm (OD₆₀₀) of 0.6 by the addition of galactose to a final concentration of 2% (w/v). After 24 h incubation at 30 °C, the cells were collected, resuspended in buffer containing 50 mM Tris–HCl pH 7.6, 1 mM EDTA, 0.6 M sorbitol and lysed by mechanical disruption as previously described³⁴. Membranes were isolated by ultracentrifugation at 4 °C and 195,000g for 2 h, homogenized in 20 mM Tris–HCl pH 7.5, 0.3 M sucrose, 0.1 mM CaCl₂, flash-frozen in liquid nitrogen and stored at –80 °C. The PfHT1-containing membranes were solubilized for 2 h at 4 °C in equilibration buffer, consisting of 1× PBS, 150 mM NaCl, 10% (v/v) glycerol and 1% (w/v) *n*-dodecyl- β -D-maltopyranoside (DDM; Glycon). Non-solubilized membranes were removed by ultracentrifugation at 195,000g for 45 min, and the cleared supernatant was incubated with 15 ml of Ni²⁺-nitrilotriacetate affinity resin (Ni-NTA; Qiagen) for 2 h at 4 °C in the presence of 40 mM imidazole under mild agitation. The resin was transferred to a 30-ml Eco-column (Bio-Rad) and washed with 300 ml of equilibration buffer containing 0.1% (w/v) *n*-undecyl- β -D-maltopyranoside (UDM; Anatrace) and 50 mM imidazole. The immobilized protein was eluted in 30 ml of equilibration buffer containing 0.1% (w/v) UDM and 250 mM imidazole. The eluate was incubated with equimolar TEV protease at 4 °C overnight to cleave the GFP–His₈ tag during dialysis performed against 3 l of dialysis buffer, consisting of 20 mM Tris–HCl pH 7.5, 150 mM NaCl and 0.08% (w/v) UDM. The dialysed and digested sample was loaded onto a 5-ml HisTrap column (GE Healthcare) equilibrated with dialysis buffer, and the PfHT1-containing flow-through was collected and concentrated. The concentrated solution was applied onto a PD-10 desalting column (Sephadex G-25, GE) pre-equilibrated in dialysis buffer, and the initial 1.6 ml of the flow-through collected, concentrated to 6–8 mg ml^{–1} and used for crystallization experiments. For proteoliposome-based transport assays, the on-column immobilized PfHT1 was washed, eluted and dialysed in equilibration buffer and dialysis buffer containing DDM 0.1% (w/v) and 0.03% (w/v), respectively, and concentrated to 2 mg ml^{–1}.

PfHT1 mutants were generated by overlap PCR, cloned into the pDDGFP₂ vector, and overexpressed in 6-l cultures as previously

described for the wild type. The PfHT1 mutants were purified as described for the wild type, but without GFP–His₈ tag removal by TEV protease cleavage. The purified PfHT1–GFP fusions were concentrated to 2 mg ml^{–1} and judged to be monodisperse by size-exclusion chromatography using an Enrich 650 10 × 300 column in buffer containing 20 mM Tris–HCl pH 7.5, 150 mM NaCl and 0.03% (w/v) DDM.

Transport activity of PfHT1 reconstituted into liposomes

Total bovine brain lipid extracts (Sigma Aldrich) and cholesteryl-hemisuccinate (CHS) (Sigma-Aldrich) powder were mixed in buffer containing 10 mM Tris–HCl pH 7.5 and 2 mM MgSO₄ to a final concentration of 30 and 6 mg ml^{–1}, respectively. The lipid mixture was subjected to multiple rounds of freeze–thaw cycles by flash-freezing in liquid nitrogen and thawing at room temperature interspersed with sonication. Lipid mixture was further spun down at 16,000g for 15 min and the supernatant containing small unilamellar vesicles was collected. To make proteoliposomes, 10 μ g of purified PfHT1 was added to 500 μ l of unilamellar vesicles, flash-frozen and thawed at room temperature. Large unilamellar proteoliposomes were prepared by extrusion (LiposoFast, Avestin; membrane pore size, 400 nm). Transport assays for PfHT1 mutants were carried as GFP fusions and compared with PfHT1 wild type prepared in the same manner.

For the transport time-course experiments, 15 μ l of prepared proteoliposomes were diluted into 45 μ l of external buffer consisting of 10 mM Tris–HCl 7.5, 2 mM MgSO₄ and either: [¹⁴C]D-glucose (30 μ M) (American Radiolabelled Chemicals and Moravék Biochemicals), [³H]D-xylose (0.3 μ M) (American Radiolabelled Chemicals), [¹⁴C]D-mannose (30 μ M) (Moravék Biochemicals) or [¹⁴C]D-galactose (30 μ M) (American Radiolabelled Chemicals), [¹⁴C]D-fructose (6.0 μ M), (Moravék Biochemicals), [³H]D-glucosamine (0.3 μ M) (Perkin Elmer). The reaction was stopped by the addition of 1 ml of 10 mM Tris–MgSO₄ buffer and followed by rapid filtering through a 0.22- μ m filter (Millipore). The on-filter collected proteoliposomes were washed with 6 ml of buffer containing 10 mM Tris–HCl 7.5 and 2 mM MgSO₄, transferred to scintillation vials and emulsified in 5 ml of Ultima Gold scintillation liquid (Perkin Elmer) before scintillation counting (TRI-CARB 4810TR 110 V; Perkin Elmer).

The proteoliposomes for [¹⁴C]D-glucose competitive-uptake assays were prepared as described for the time-course experiments. [¹⁴C]D-glucose competitive uptake was measured at 30 s in external buffer containing unlabelled sugars at a final concentration of 50 mM.

For kinetic analysis, the K_M and V_{max} values for D-glucose and D-fructose transport for PfHT1 and mutants were determined by measuring the initial transport velocities for D-glucose at 20 s and D-fructose at 60 s for increasing concentrations of these sugars in buffer containing stoichiometric amounts of [¹⁴C]D-glucose or [¹⁴C]D-fructose. The recorded radioactivity from empty liposomes was subtracted from the recorded decay counts of the transported sugars and fitted to Michaelis–Menten kinetics using nonlinear regression by GraphPad Prism 7.0. K_M and V_{max} values for D-mannose, D-galactose and D-glucosamine for PfHT1 were determined in the same way, with initial transport velocities recorded at 20 s for D-mannose and 60 s for D-galactose and D-glucosamine. Time course and kinetics of PfHT1 and PfHT1–GFP were measured to be comparable. To calculate k_{cat} , the amount of transported sugar was therefore normalized by the fraction of reconstituted PfHT1–GFP fusion incorporated into liposomes, which could be calculated by fluorescence-detection size-exclusion chromatography (FSEC)³⁵ analysis of 3% DDM (w/v)-solubilized proteoliposomes. PfHT1–GFP orientation into liposomes was estimated by incubating 150 μ l of proteoliposomes with and without TEV protease at a ratio of 1:3 (w/w) overnight at 4 °C and the fraction of cleaved GFP estimated by in-gel fluorescence³⁴ with a ratio of about 60:40 (outside: inside) calculated.

For inhibition assays, 15 μ l of PfHT1 wild type- or Trp412Ala mutant-containing proteoliposomes were diluted into 43 μ l of buffer consisting of 10 mM Tris–HCl 7.5, 2 mM Tris–MgSO₄ that had been pre-incubated for 1 h in either 4% (v/v) DMSO or 4% (v/v) DMSO with 115 μ l of the tested

compounds for inhibition; C3361 was synthesized by BOC Sciences and the MVV009085-homologue was supplied by Mcule. Transport was initiated by the addition of 2 μ l of [14 C]-D-glucose at 40 μ M final concentration. The reaction was stopped by the addition of 1 ml of 10 mM Tris-MgSO₄ buffer and followed by rapid filtering through a 0.22- μ m filter (Millipore). The on-filter collected proteoliposomes were washed with 6 ml of buffer containing 10 mM Tris-HCl 7.5 and 2 mM MgSO₄, transferred to scintillation vials and emulsified in 5 ml of Ultima Gold scintillation liquid (Perkin Elmer) before scintillation counting (TRI-CARB 4810TR 110 V; Perkin Elmer).

The recorded radioactivity of empty liposomes at 4% (v/v) DMSO was subtracted from each tested condition. Half-maximal inhibitory concentration (IC₅₀) values were obtained by fitting one-phase decay nonlinear regression by GraphPad Prism 7.0. All transport results are represented as mean values ($n = 3$) with their corresponding standard errors.

Crystallization and structure determination of PfHT1

Crystals of PfHT1 in complex with D-glucose were grown at 4 °C using the hanging-drop vapour-diffusion method. Purified PfHT1 protein at 8 mg/ml was added D-glucose to a final concentration of 50 mM. One microlitre of this solution was mixed 1:1 with reservoir solution consisting of 0.1 M MES pH 6.5, 0.1 M MgCl₂, 26–30% (w/v) PEG 300 and 0.2% (w/v) *n*-nonyl- β -D-glucopyranoside (NG, Anatrace). Crystals appeared within 1 week in 26% PEG 300 and were dehydrated by equilibration of the drops against 500 μ l reservoir solution containing increasing concentrations of PEG 300 in steps of 2% (w/v) up to a final concentration of 32% (w/v). Crystals were subsequently collected and flash-frozen in and stored under liquid nitrogen.

X-ray diffraction data from PfHT1 crystals were collected at 100K at the European Synchrotron Radiation Facility (ESRF) at the beamlines ID30A-3 and ID23-1. Two datasets from different crystals were indexed, integrated and scaled together using XDS³⁶ before merging using Aimless³⁷. Initial phases of PfHT1 were obtained by molecular replacement using phenix.mr_rosetta^{38,39} and the outward-facing occluded structure of human GLUT3 as an input search model (PDB 4ZW9). There are four PfHT1 molecules in the asymmetric unit. Structure refinement was carried out using Phenix.refine^{40,41} and auto BUSTER⁴² with local NCS (non-crystallographic symmetry), one TLS (translation–libration–screw rotation) group per chain and external constraints to human GLUT3, interspersed with manual model building in Coot⁴³. The Ramachandran statistics are 92.6% favoured, 6.91% allowed and 0.93% outliers. Other data collection and refinement statistics are presented in Extended Data Table 2. Structural alignments were performed using the align command of PyMol software (<http://www.pymol.org/>) using C α coordinates.

Molecular dynamics simulations

The starting models used for molecular dynamics simulations were human GLUT3 (PDB 4ZW9) and chain C of PfHT1. The cytosolic loops of PfHT1 connecting TM5 to TM6 and TM9 to TM10, as well as ICH5, were modelled using MODELLER⁴⁴ version 9.21 before simulations. Six simulation systems were constructed (three for PfHT1 and three for GLUT3), each of which consisted of the protein embedded in a POPC bilayer. To do this, six lipid configurations were generated using the CHARMM-GUI membrane builder⁴⁵, in which the protein (and ligands if applicable) was embedded. These systems were then solvated in 150 mM NaCl. Details of each simulation replica can be found in Supplementary Table 1.

All systems underwent energy minimization using steepest descent. Equilibration molecular dynamics was then performed for a total of 375 ps, gradually relaxing positional restraints on protein, POPC lipids and ligands, when relevant. The duration of each production molecular dynamics simulation can be found in Supplementary Table 1. Simulations were carried out under periodic boundary conditions and

production molecular dynamics was carried out using a 2-fs time steps. The temperature and pressure were maintained at 303.15K and 1 bar using the Berendsen thermostat and barostat⁴⁶, respectively. Pressure coupling was performed using semi-isotropic coupling with a time constant of 5 ps and compressibility of 4.5×10^{-5} bar⁻¹. Temperature coupling was performed using three separate groups for protein, lipids and solvent. Hydrogen bonds were constrained using the linear constraint solver (LINCS)⁴⁷. Electrostatic interactions were modelled with a 1.2-nm cutoff, with a switching function between 1.0 and 1.2 nm. Long-range electrostatics were calculated using particle mesh Ewald (PME)⁴⁸. All-atom molecular dynamics simulations were performed using Gromacs 2018.1. Interactions were modelled with the CHARMM36m (protein, lipids and ions) and the TIP3P (water) forcefields⁴⁹.

Analysis of simulations and protein morphing

Analysis of the molecular dynamics simulations were performed using the gromacs analysis tools gmx rmsf and gmx pairdist, for root mean square fluctuation and gating-residue distance calculations. Gate distance was determined by measuring the centre of mass of residues that remained closest around the extracellular gate during the simulations: residues Val44 to Ile50 and Asn311 to Ser317 for PfHT1, and residues Thr28 to Pro34 and Asn286 to Ser292 for human GLUT3. The gate distances plotted are the mean between the three replicas are indicated by darkened lines (Fig. 3e). Python scripts were written to parse and plot relevant data⁵⁰. Figure generation was performed using PyMol (<https://pymol.org/2/>).

Morphing between structural states for movie generation was performed using PyMol. From the PfHT1 structure presented here, chain C was used to generate models in the following resolved conformations: outward-open (GLUT5, PDB 4YB9), outward-occluded (GLUT3, PDB 4ZW9), inward-occluded (Xyle, PDB 4JA3) and inward-open (GLUT5, PDB 4YBQ). Structural alignments of the proteins were carried out in PyMol, and subsequent model generation using MODELLER. The N termini of these respective models were superimposed, morphed between states and the video was made using PyMol.

Sugar-porter principal component analysis

Principal component analysis (PCA) is a statistical technique to reveal dominant patterns⁵¹. Diagonalization of the covariance matrix of a system of variables renders the major axes of statistical variance or principal components, thus mapping complex multidimensional data into a few coordinates, which contain the major trends that explain the statistical variation. For the sugar-porter structures, a set of near-intact structures sharing 30% homology with PfHT1—that is, eukaryotic GLUT structures (GLUT1, GLUT3 and GLUT5) along with *E. coli* Xyle transporter (16 PDB codes in total: PfHT1, 6RW3; Xyle, 4GBZ, 4GBY, 6N3I, 4JA3 and 4JA4; GLUT1, 4PYP, 5EQI, 5EQH and 5EQG; GLUT3, 5C65, 4ZWC, 4ZWB and 4ZW9; and GLUT5, 4YB9 and 4YBQ)—were aligned to extract the common structural fold, mostly formed by conserved helices (353 residues) (Extended Data Fig. 9a). Missing residues in some of the structures were rebuilt with MODELLER, making sure that the positions of the corresponding structural elements were strictly kept for the core alignment. The structural ensemble was aligned to the structure of GLUT5 in an open outward-facing conformation (root mean square deviation (r.m.s.d.) of 2.7 ± 1.2 Å) and used to compute the covariance matrix versus this reference; that is, the mean-square deviations in atomic coordinates from their mean position (diagonal elements) and the correlations between their pairwise fluctuations (off-diagonal elements). The covariance matrix was diagonalized to obtain a set of eigenvectors or principal components, ordered according to their eigenvalues with decreasing variance from those representing the largest-scale motions up to the smallest fluctuations in atomic coordinates. Within this framework, any structure *i* is characterized by its scalar product projections onto the conformational space defined by the major components, \mathbf{PC}_k ($k = 1, 2, \dots, n$):

$$\mathbf{PC}_k = \mathbf{T}_{i-0} \cos(\mathbf{PC}_k \wedge \mathbf{T}_{i-0})$$

in which \mathbf{T}_{i-0} is the vector between the coordinates of i structure and the chosen reference 0 (4YBQ, in this case), and \mathbf{PC}_k is one of the major principal component axes, which can classify and cluster structures, and extract motion information and transition pathways from them^{52,53}. For the MFS ensemble, the first component alone captures about 65% of the structural variation associated with the rocker switch, thus separating the crystallographic structures along the transport cycle. The angle between the sugar-porter tandem repeats was estimated as the angle formed by TM2 and TM8, using an in-house Visual Molecular Dynamics script.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The coordinates and the structure factors for PfHT1 have been deposited in the PDB 6RW3. All data are available in the paper or Supplementary Information.

33. Kota, J., Gilstring, C. F. & Ljungdahl, P. O. Membrane chaperone Shr3 assists in folding amino acid permeases preventing precocious ERAD. *J. Cell Biol.* **176**, 617–628 (2007).
34. Drew, D. et al. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat. Protoc.* **3**, 784–798 (2008).
35. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
36. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
37. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
38. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
39. DiMaio, F. et al. Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **10**, 1102–1104 (2013).
40. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
41. Headd, J. J. et al. Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D* **68**, 381–390 (2012).
42. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68**, 368–380 (2012).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
44. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
45. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
46. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684 (1984).
47. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
48. Darden, T., Darrin, Y. & Pedersen, L. Particle mesh Ewald: An N -log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
49. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
50. van der Walt, S. Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22 (2011).
51. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
52. Orellana, L., Yoluk, O., Carrillo, O., Orozco, M. & Lindahl, E. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat. Commun.* **7**, 12575 (2016).
53. Orellana, L., Gustavsson, J., Bergh, C., Yoluk, O. & Lindahl, E. eBDIMS server: protein transition pathways with ensemble analysis in 2D-motion spaces. *Bioinformatics* **35**, 3505–3507 (2019).

Acknowledgements We thank D. Daley for advice on PfHT1 overexpression optimization, and G. von Heijne and S. Newstead for critical reading of the manuscript. X-ray diffraction data were collected at the European Synchrotron Radiation Facility beamlines, the Diamond Light Source beamlines and the MaxIV BioMax beamline with assistance from beamline scientists. This work was funded by the Knut and Alice Wallenberg Foundation (D.D.) and the Science for Life Laboratory (L.D.); D.D. acknowledges support from EMBO through the Young Investigator Program (YIP).

Author contributions D.D. designed the project. Cloning, expression screening and crystallization of PfHT1 were carried out by A.A.Q. Data collection, structure determination and refinement of PfHT1 were carried out by A.A.Q., E.N., J.B., R.M., M.C. and D.D. Experiments for functional analysis were carried out by A.A.Q. and A.S. Molecular dynamics simulations of PfHT1 were carried out by S.E.M. and L.D. Sugar porter PCA was carried out by L.O. The manuscript was prepared by D.D. with contributions from all authors.

Competing interests The authors declare no competing interests.

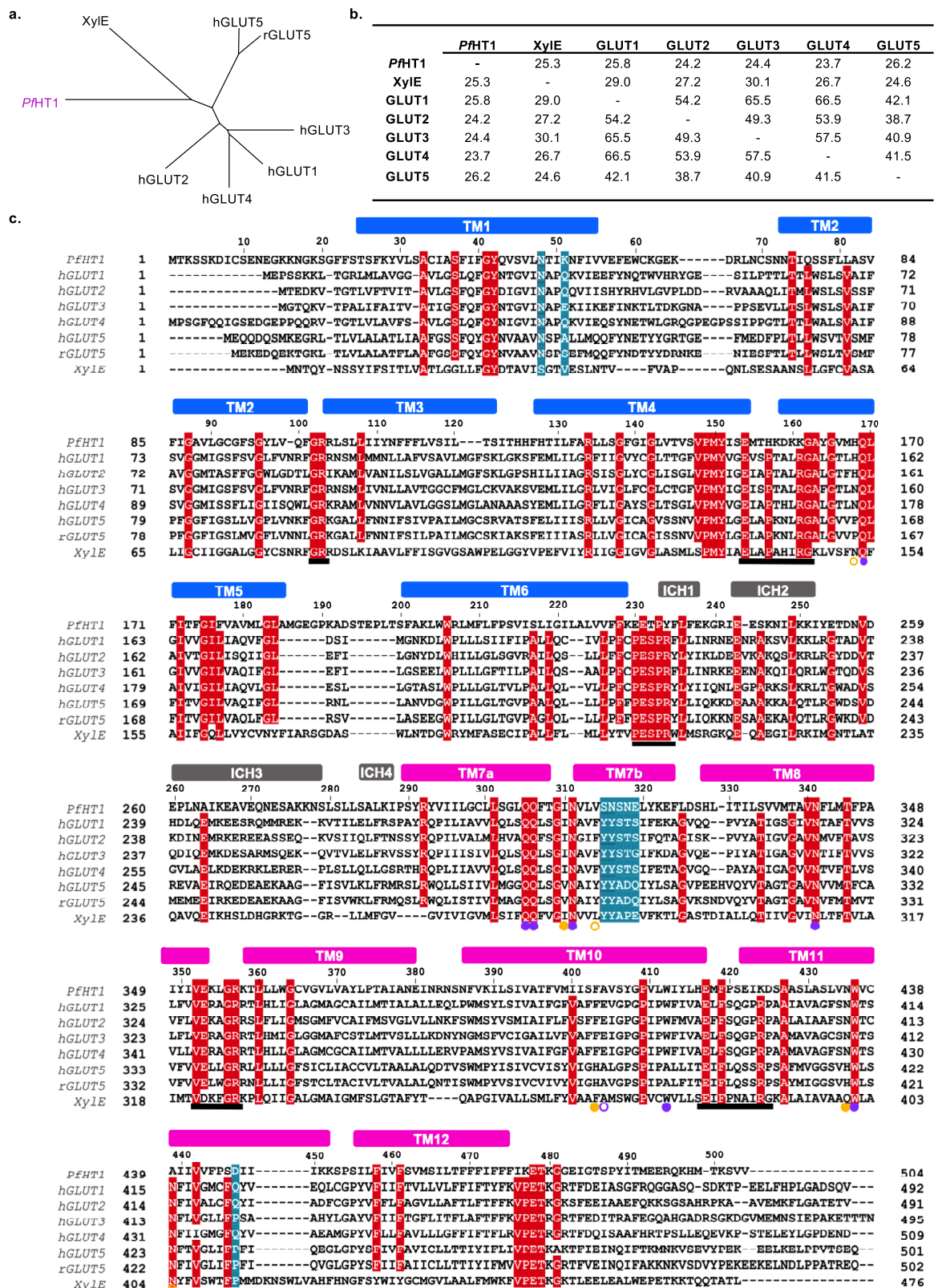
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1963-z>.

Correspondence and requests for materials should be addressed to D.D.

Peer review information Nature thanks Jeff Abramson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

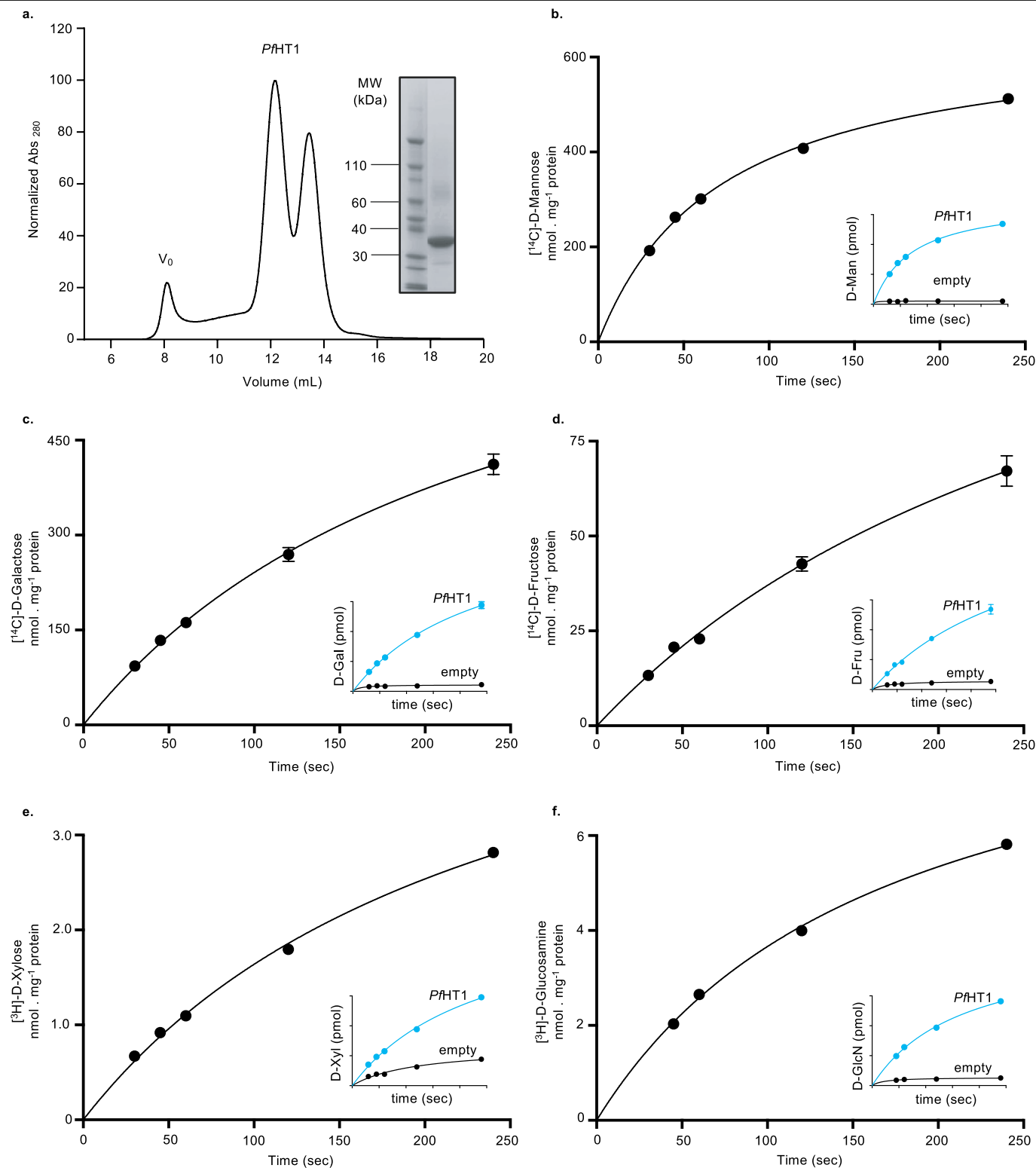
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

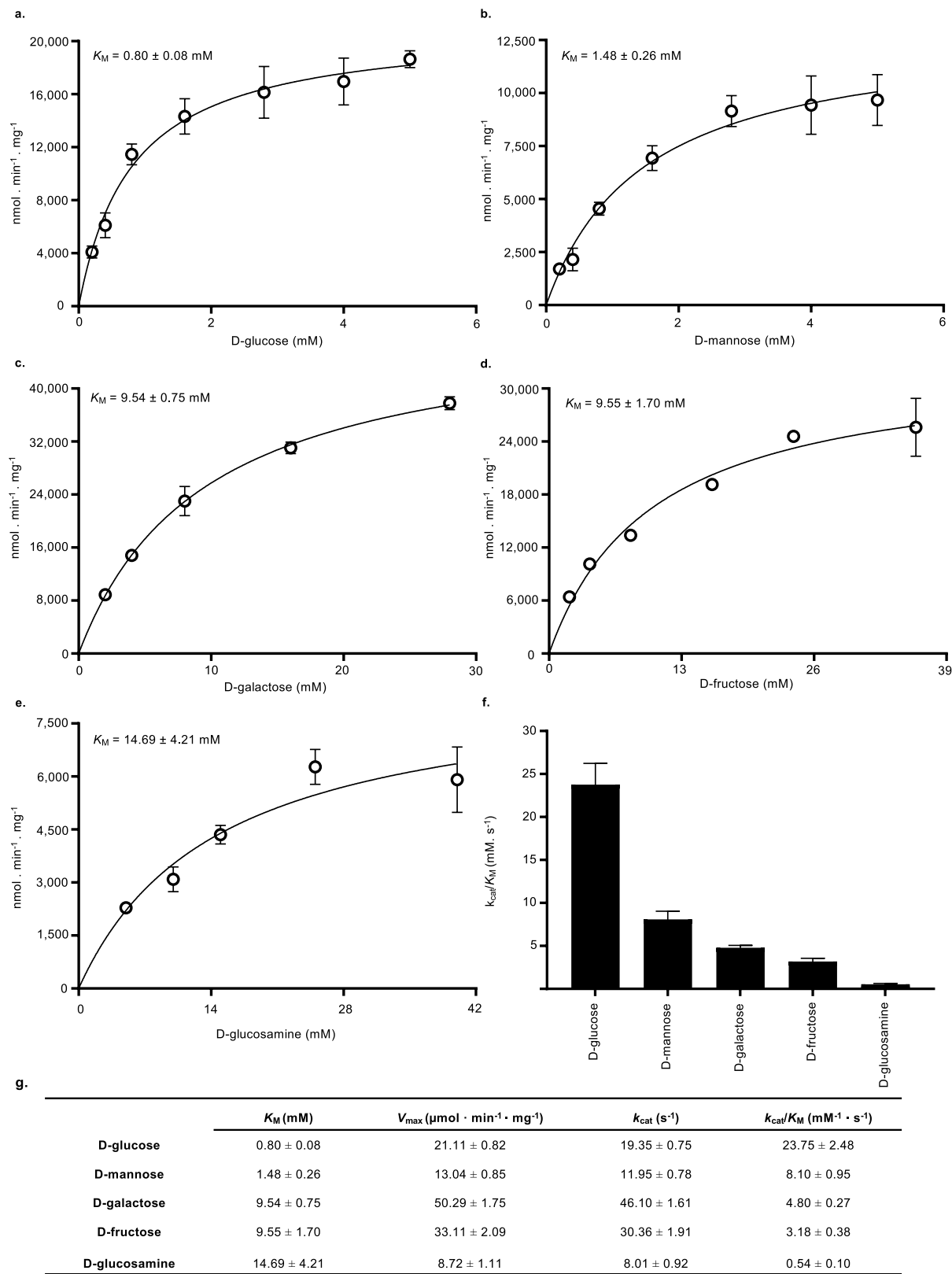
Extended Data Fig. 1|PfHT1 is distantly related to both the human GLUT transporters and the bacterial homologue Xyle. **a**, Unrooted phylogenetic tree of human GLUT1, GLUT2, GLUT3, GLUT4 and GLUT5 (GLUT1–5), rat GLUT5, *E. coli* Xyle and PfHT1 **b**, Table of protein sequence identity of proteins shown in **a**. (only rat (and not human) GLUT5 is shown). **c**, Sequence alignment of PfHT1, human GLUT1–5, rat GLUT5 and *E. coli* Xyle. Secondary structure elements of PfHT1 are indicated above the alignment, and coloured as in Fig. 1c. Residues conserved in at least 80% of the alignment are highlighted by red boxes, and gating residues are highlighted by blue boxes. Conserved binding-site residues between PfHT1 and human GLUT3 are indicated with purple filled

dots and non-conserved residues are indicated with non-filled purple dots. Conserved residues close to the binding site are indicated with yellow filled dots and non-conserved residues with non-filled yellow dots. Black bars beneath the alignment indicate residues in the sugar-porter motifs^{15,16}. The Uniprot reference numbers of the alignment proteins are: PfHT1 (Q7KWJ5), human GLUT1 (P11166), human GLUT2 (Q102R8), human GLUT3 (P11169), human GLUT4 (P14672), human GLUT5 (P22732), rat GLUT5 (P43427) and Xyle (P0AGF4). For the sake of clarity, residues 109–121 from Xyle and residues 54–86 from human GLUT2 were omitted.



Extended Data Fig. 2 | PfHT1 has evolved to be an efficient polyspecific sugar transporter. **a**, Size-exclusion chromatogram of DDM-purified PfHT1 showing PfHT1 migrates as two oligomeric species (dimer and monomer); the sample migrates as a monomer during SDS-PAGE. **b**, Time-dependent uptake of [¹⁴C] D-mannose (black circles) by PfHT1 in proteoliposomes. Inset, PfHT1 uptake of

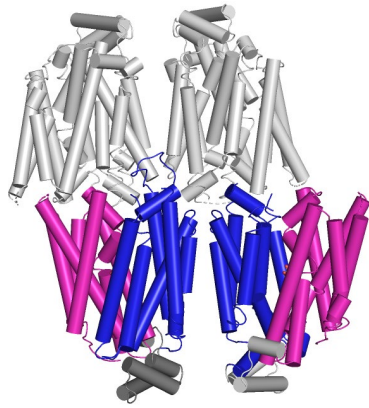
radiolabelled sugar (cyan trace) compared with non-specific uptake estimated from radioactivity measured from liposomes incubated without protein (black trace). Error bars represent the mean \pm s.e.m. of $n = 3$ biologically independent experiments. **c**, As in **b**, for [¹⁴C]D-galactose. **d**, As in **b**, for [¹⁴C]D-fructose. **e**, As in **b**, for [³H]D-xylose. **f**, As in **b**, for [³H]D-glucosamine.



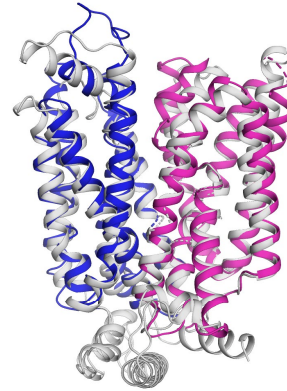
Extended Data Fig. 3 | PfHT1 has evolved to be an efficient polyspecific sugar transporter. **a**, Zero *trans* kinetics of PfHT1 D-glucose transport. Kinetic curves were fitted from data points recorded at increasing D-glucose concentrations after 20 s and fitted by nonlinear regression. Error bars represent mean \pm s.e.m. of $n = 3$ biologically independent experiments. **b**. As in **a**, for D-mannose. **c**. As in **a**, for D-galactose and except that time points were recorded after 60 s. **d**. As in **a**, for D-fructose and except that time points were

recorded after 60 s. **e**. As in **a**, for D-glucosamine. **f**. Bars represent specificity constant (k_{cat}/K_M) values of PfHT1 for different sugars as tabulated and described in **g**. **g**. The fitted values reported for the Michaelis constant (K_M) and V_{max} of PfHT1 for different transported sugars are mean \pm s.e.m. of $n = 3$ biologically independent experiments. Turnover (k_{cat}) and specificity constant (k_{cat}/K_M) values are derived from these kinetic parameters and adjusted to the amount of protein reconstituted into liposomes (Methods).

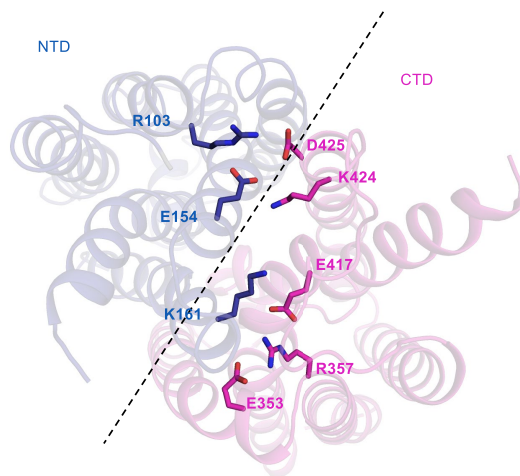
a.



b.

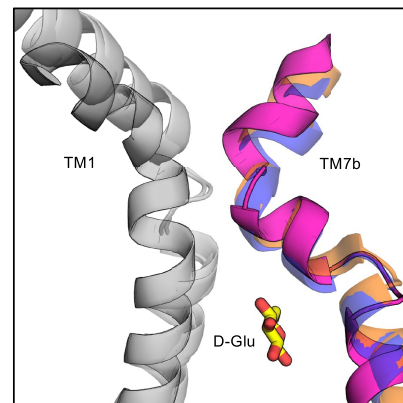
*Pfl*HT1 vs GLUT3 (grey)

c.

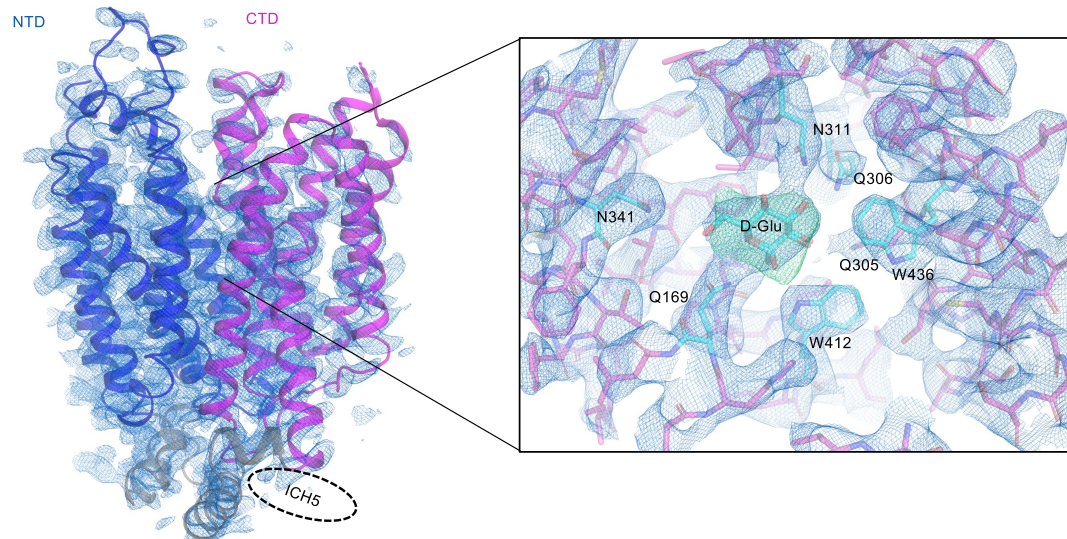
*Pfl*HT1 intracellular salt-bridge network

d.

occluded *Pfl*HT1
 Inward-open GLUT5
 Inward-open GLUT1



e.



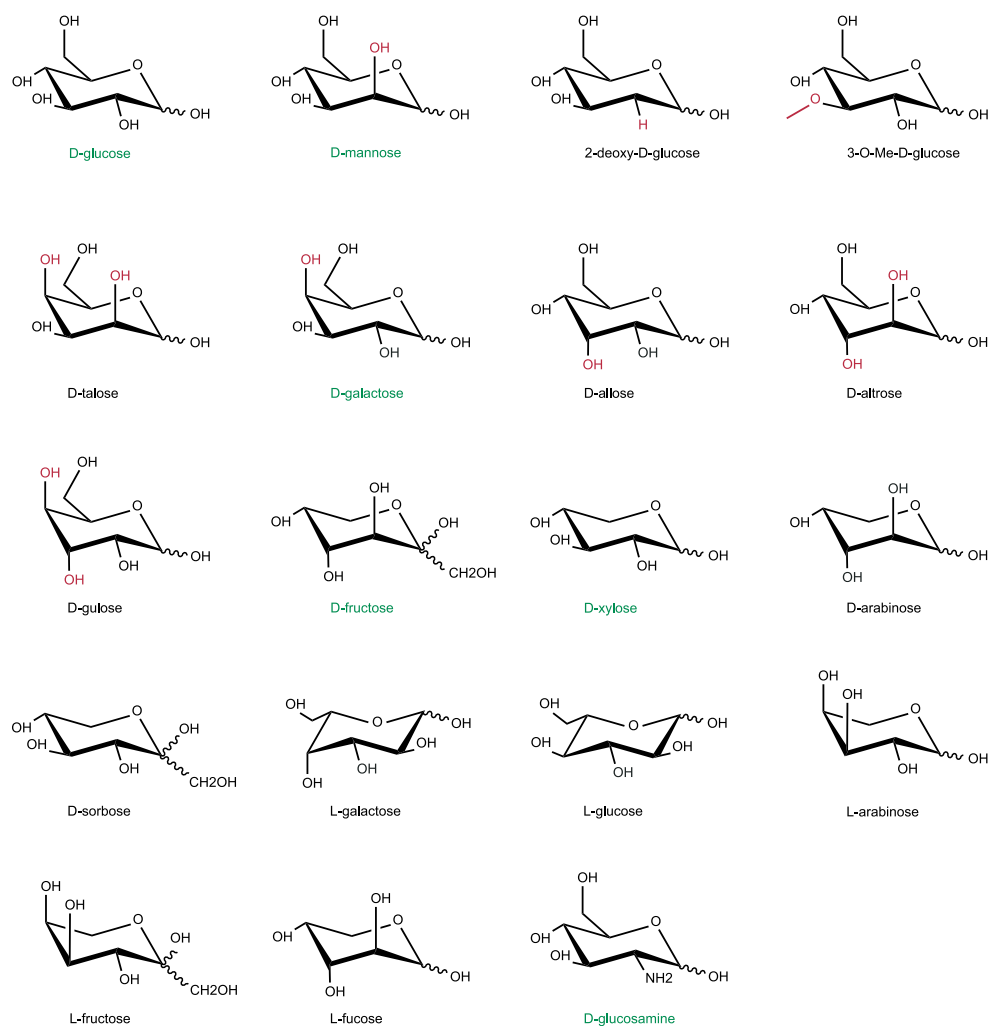
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Overall structural features of the PfHT1 structure.

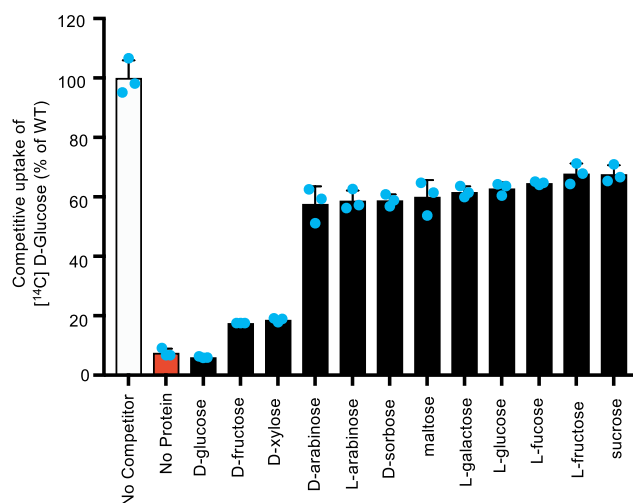
a, PfHT1 crystallized as a dimer with four molecules in the asymmetric unit. The shared dimer interface was formed between the respective N-terminal domains (blue) that—although not extensive (522 Å buried surface area)—are consistent with the fact that a fraction of purified PfHT1 migrates as a dimer by size-exclusion chromatography (Extended Data Fig. 2a). Notably, the gating helix TM7b is not making any crystal contacts. **b**, Superposition of the outward-occluded GLUT3 (PDB 4ZWB) (grey) and the occluded PfHT1 structures. The r.m.s.d. is 1.4 Å for 446 pairs of Cα atoms (Methods). **c**, Cartoon representation of PfHT1 as viewed from the cytoplasm. Blue, NTD; magenta, CTD. ICHs are not

shown for clarity. Interdomain salt-bridge-forming residues are shown as sticks, and labelled. **d**, Cartoon representation of TM7b of human GLUT1 (PDB 4PYP) (orange) and bovine GLUT5 (PDB 4YB9) (purple) in the inward-open conformation, and PfHT1 in the D-glucose-bound (yellow sticks) occluded conformation (magenta). **e**, Electron density map $2F_o - F_c$ (1.5σ) (blue mesh) for the PfHT1 structure (left) and the D-glucose residues in the sugar-binding pocket in cyan (right). The $F_o - F_c$ (3.0σ) (green mesh) maps before addition of and refinement in the presence of D-glucose are also shown. Despite the high quality of the maps, we observed no electron density for ICH5 (location in human GLUT3 shown as a dashed ellipse).

a.

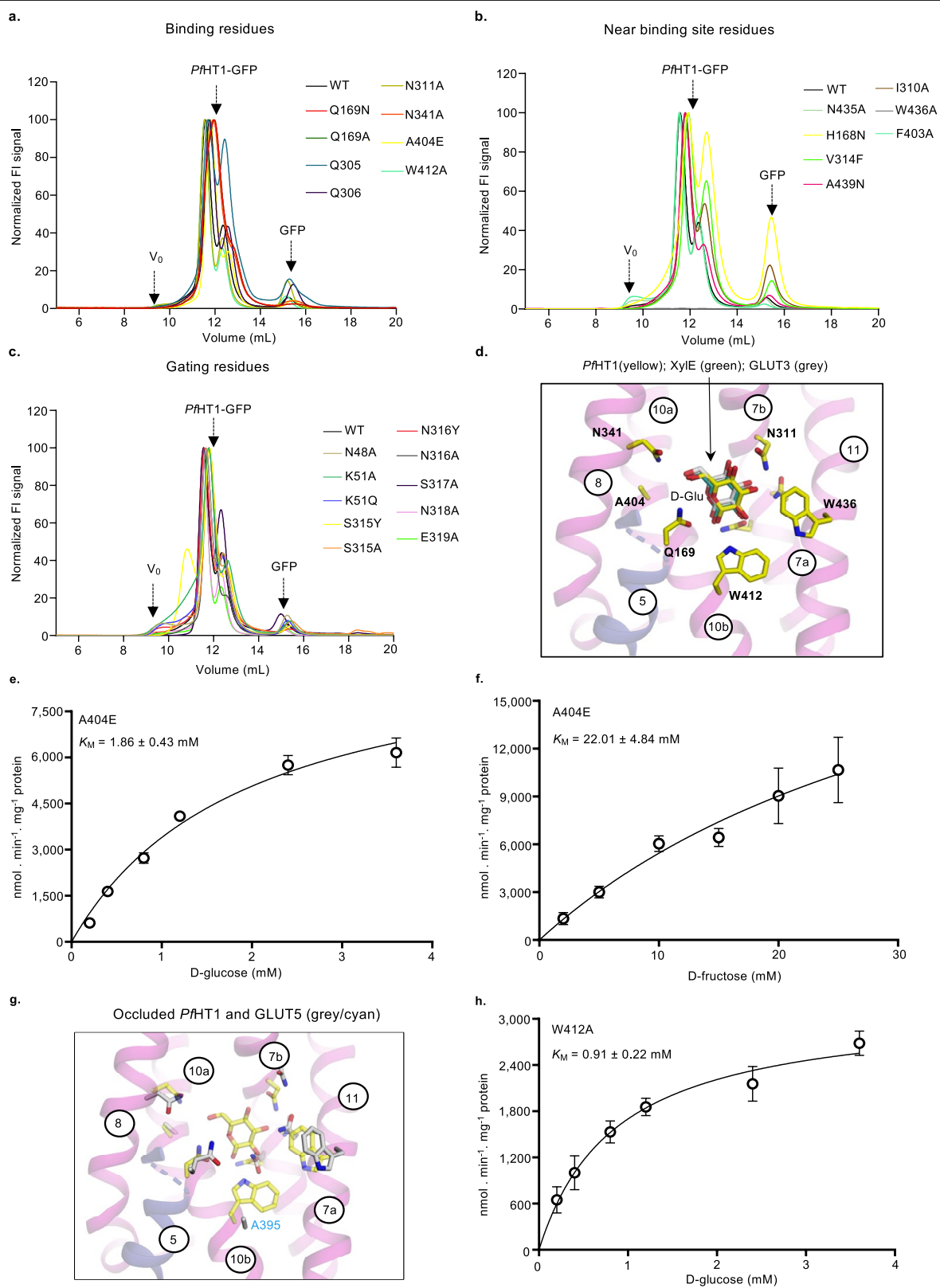


b.



Extended Data Fig. 5 | Sugars assessed for inhibition of PfHT1. a, Chemical structures of the investigated sugars used in competitive-uptake assays. Differences in hydroxyl-group position of the respective D-glucose epimers are coloured red. Sugars labelled green were also tested as radiolabelled substrates in time-course experiments (Fig. 1b, Extended Data Fig. 2b–f). **b,** The

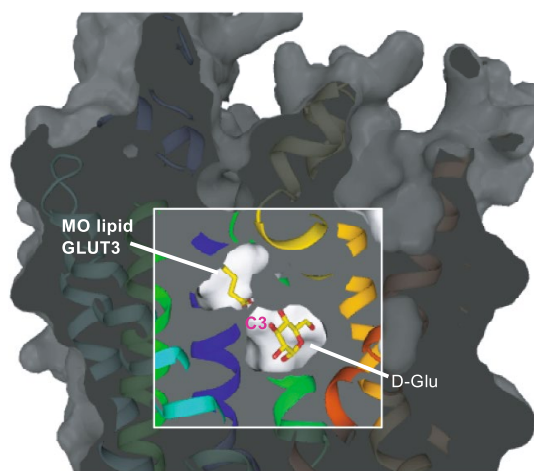
competitive uptake of $[^{14}\text{C}]$ D-glucose by PfHT1 in proteoliposomes in the absence (white bars) and presence of non-labelled sugars (black bars). Non-specific uptake was estimated from radioactivity measured from liposomes incubated without protein (red bar). Error bars represent mean and s.e.m. of $n = 3$ biologically independent experiments.



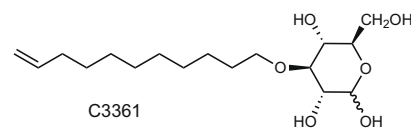
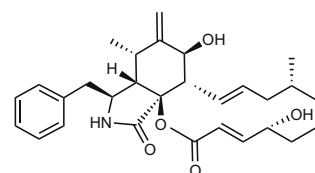
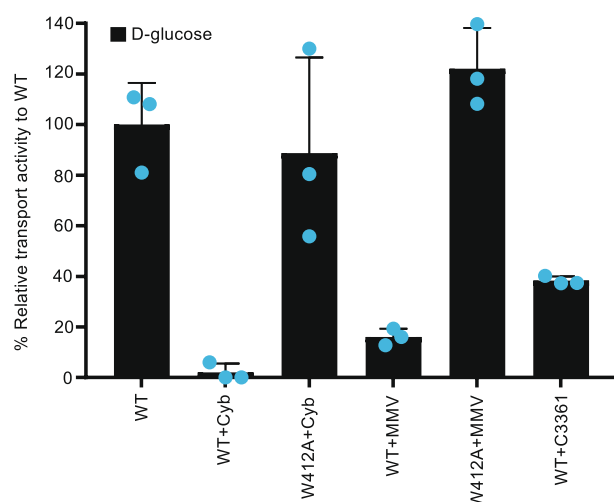
Extended Data Fig. 6 | Assessed quality of purified PfHT1–GFP fusions and analysis of the sugar-binding pocket of PfHT1. **a**, FSEC traces of DDM-purified PfHT1–GFP wild type and mutants of the sugar-binding pocket; PfHT1–GFP migrates as two species (dimer and monomer), consistent with purified PfHT1 (Extended Data Fig. 2a). FSEC traces were recorded at least twice for wild type and each respective mutant. **b**, As in **a**, for mutants peripheral to the sugar-binding pocket. **c**, As in **a**, for mutants located in TM1 and TM7b. **d**, Cartoon representation of PfHT1 with D-glucose and interacting residues labelled, shown as yellow sticks. The position of D-glucose in *E. coli* Xyle (green) (PDB 4JA3) and D-glucose in human GLUT3 (grey) (PDB 4ZW9) are shown as sticks,

after protein superimposition. **e**, Determination of the Michaelis constant (K_M) for D-glucose by the PfHT1 mutant Ala404Glu, constructed to mimic the human GLUT3 binding site. Kinetic curves were fitted from data points recorded over a range of increasing D-glucose concentrations after 90 s, and fitted by nonlinear regression using data from $n = 3$ biologically independent experiments (values reported are mean \pm s.e.m. of the fit). **f**, As in **e**, for D-fructose. **g**, Sugar-binding-site comparison between PfHT1 side chains (yellow sticks) and rat GLUT5 side chains (conserved side chains, grey sticks; non-conserved side chains, cyan sticks). **h**, As in **e**, for the PfHT1 mutant Trp412Ala constructed to mimic the rat GLUT5 binding site.

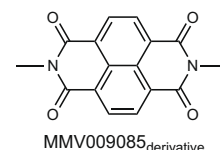
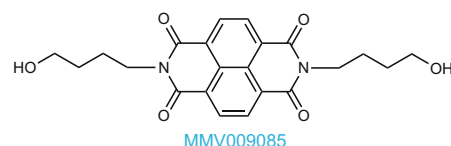
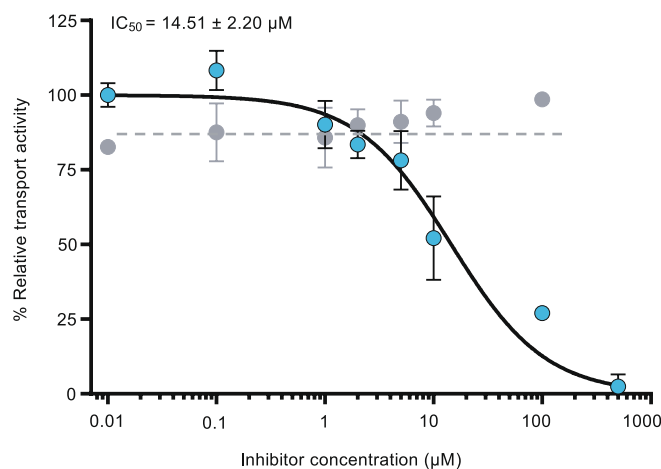
a.



b.



c.

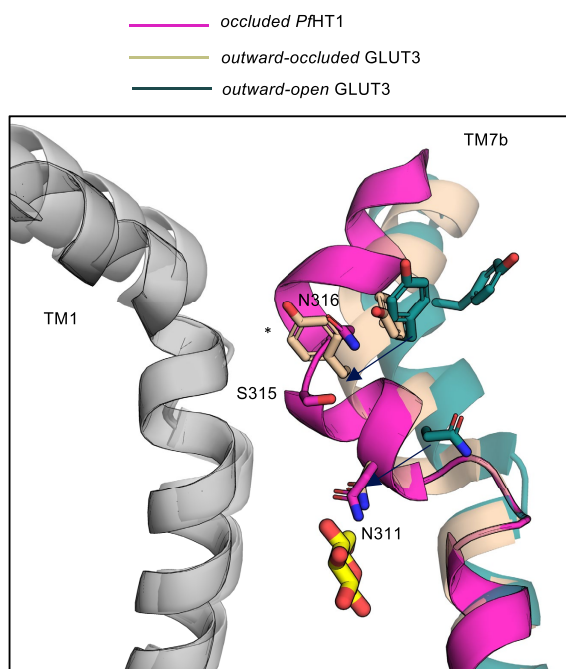


Extended Data Fig. 7 | Small-molecule-inhibition analysis of PfHT1.

a, Surface transversal cross-sections through the membrane of the PfHT1 structure in the occluded conformation with D-glucose shown as sticks, and a side vestibule accessible to the C3- and C4-hydroxyl groups that—in human GLUT3—was occupied by a monoolein lipid. **b**, The competitive uptake of [¹⁴C]D-glucose (black bars) by PfHT1 wild type and the mutant W412A in proteoliposomes in the absence and presence of cytochalasin B (Cyb) (120 μM),

MMV009085 (MMV) (100 μM) or C3361 (100 μM). Error bars represent s.e.m. of $n = 3$ biologically independent experiments (left). Cytochalasin B and the inhibitor C3361 (which is D-glucose with a undec-10-en chain at the C3-hydroxyl position) are shown (right). **c**, IC₅₀ curves for MMV009085 (blue filled circles) and derivative lacking the butanol tails (grey filled circles). Error bars represent s.e.m. of $n = 3$ biologically independent experiments (left). Structures of MMV009085 and derivative are shown on the right.

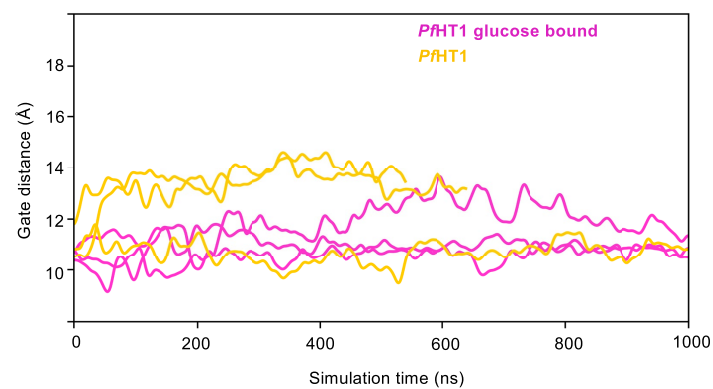
a.



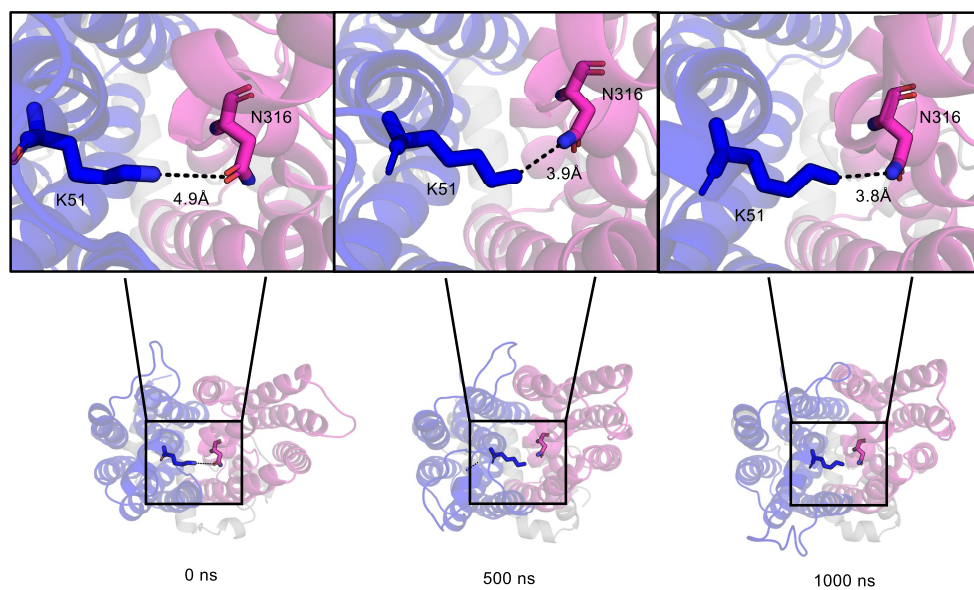
b.



c.



d.



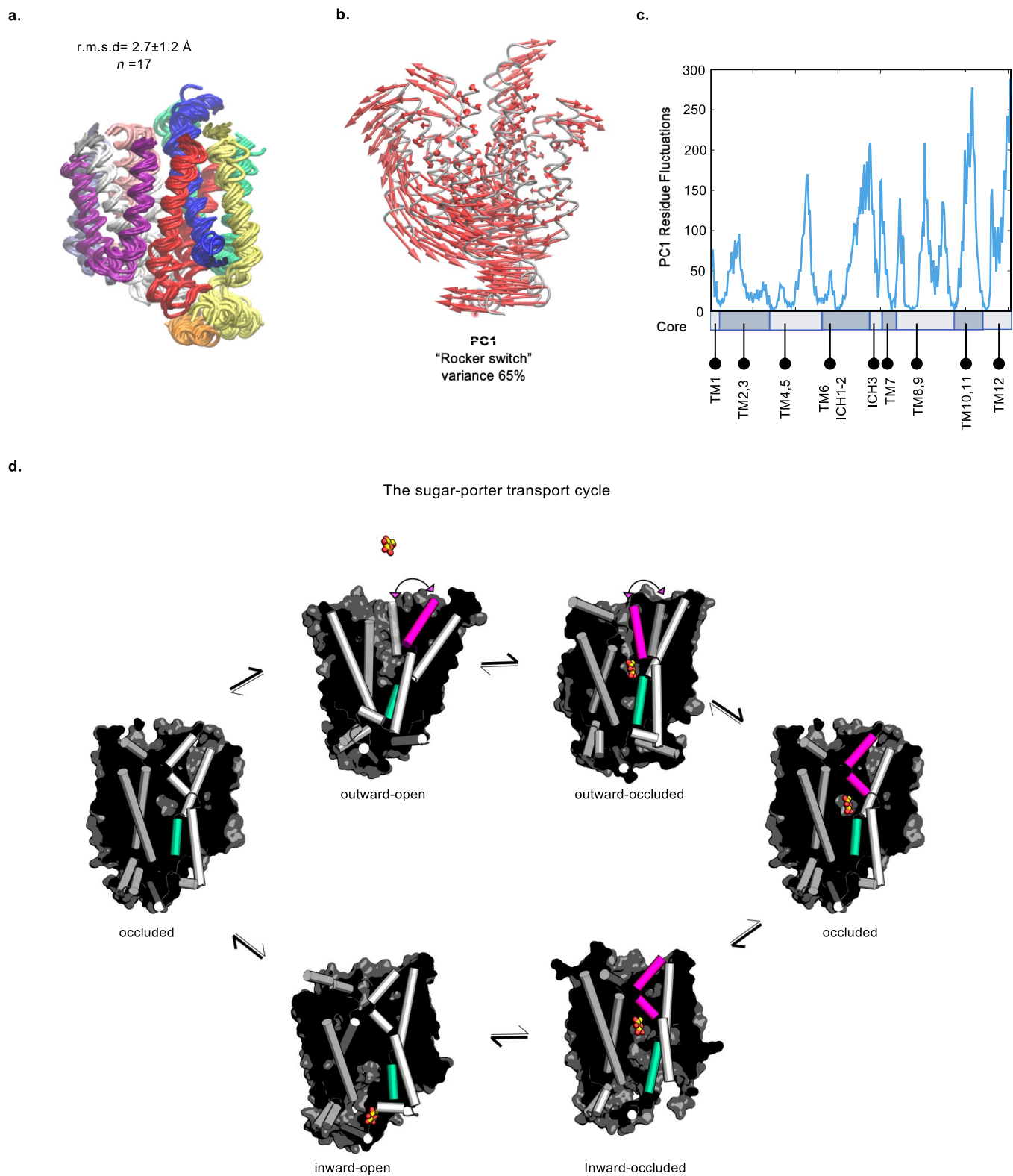
Extended Data Fig. 8 | See next page for caption.

Article

Extended Data Fig. 8 | The extracellular substrate-gating helix TM7b.

a, Cartoon representation of TM7b of human GLUT3 in the outward-open (green) (PDB 4ZWC) and outward-occluded D-glucose-bound conformation (light-brown) (PDB 4ZW9) and of PfHT1 in the D-glucose-bound occluded conformation (magenta). Arrows indicate that inward movement of TM7b is coupled to coordination of the C3- and C4- hydroxyl groups of D-glucose (shown as yellow sticks) by the strictly conserved asparagine residue corresponding to Asn311 in PfHT1. Only in PfHT1 does TM7b break into two perpendicular segments as observed in the inward-facing structures of GLUT1 and GLUT5 (Extended Data Fig. 4d). The asterisk highlights the highly conserved tyrosine residues that occlude the substrate from exiting in the outward-occluded conformations, and which—in PfHT1—are replaced by serine

(S315) and asparagine (N316). **b**, Molecular dynamics simulations of TM1–TM7 gating interactions: D-glucose-bound outward-occluded gating interactions by human GLUT3 in the presence (yellow) and absence (grey) of D-glucose. The distribution of conformations, shown as the gating distance, from three independent 1- μ s molecular dynamics simulations, as described in the Methods. **c**, As in **b**, for D-glucose-bound outward-occluded gating interactions by PfHT1 in the presence (magenta) and absence (yellow) of D-glucose. The distribution of conformations, shown as the gating distance, from three independent 1- μ s molecular dynamics simulations, as described in the Methods. **d**, Snapshots of the distance between residues Lys51 and Asn316 at 0-ns, 500-ns and 1,000-ns time points.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | PCA analysis of sugar-porter structure and alternating-access mechanism.

a, Aligned core of near-intact MFS transporter structures from 17 structures, including PfFHT1, mammalian (GLUT1, GLUT3 and GLUT5) and prokaryotic (XylE) systems. The 'core' structural elements conserved among the structures are shown in different colours (Methods).

b, Motions along the first principal component (PC1) derived from the core ensemble, tracking the rocker-switch motion. **c**, Residue fluctuations computed from PC1, with the core helix fragments shown as shadowed areas at the base. **d**, Schematic of the structural basis of the sugar-porter alternating-access mechanism. To summarize, in the outward and outward-occluded conformations (PDB 4YBQ and 4ZW9) the substrate-gating helix TM7b (magenta and transparent) is mobile and samples either state, as seen in molecular dynamics simulation of human GLUT3; spontaneous gate closure is further consistent with the fact that—even in the presence of maltose—GLUT3 crystallizes in both outward-open and outward-occluded conformations⁴. Substrate binding conformationally stabilizes the outward-occluded state, thus increasing the likelihood for TM7b to break in the middle, completely

close the substrate pocket and form contacts with TM1. In the occluded state, the salt-bridge interactions between ICH5 in the C-terminal bundle and ICH1, ICH2 ICH3 and ICH4 are lost, which indirectly destabilizes the highly conserved intrabundle salt-bridge network. Breakage of the intrabundle salt-bridge network catalyses global rocker-switch rearrangements of the N- and C-terminal bundles. In the inward-occluded conformation (PDB 4JA3), the intracellular gating helix TM10b (cyan)—which is related by inverted symmetry to TM7b—spontaneously moves outward to the inward-open conformation (PDB 4YB9). After sugar release, the sugar porter spontaneously resets itself to the outward-facing conformation through an 'empty' occluded state²². Spontaneous resetting means that the energetic barriers separating opposite-facing states must be low enough that the occluded state can form in the absence of sugar binding. Nevertheless, consistent with a conformational-selection-driven rocker-switch mechanism, substrate binding catalyses transport as rates are substantially faster through 'substrate-bound' versus 'empty' occluded-state transitions¹.

Extended Data Table 1 | Zero trans proteoliposome (lacking internal sugar) in vitro kinetics of D-glucose and D-fructose by PfHT1 wild type and mutants

	K_M (mM)		V_{max} ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)		k_{cat} (s^{-1})		k_{cat}/K_M ($\text{mM}^{-1}\cdot\text{s}^{-1}$)		$ratio\ k_{cat}/K_M$
	glucose	fructose	glucose	fructose	glucose	fructose	glucose	fructose	glucose/fructose
PfHT1 WT	0.80 ± 0.10	9.51 ± 1.70	21.12 ± 0.80	33.12 ± 2.14	19.32 ± 0.60	30.36 ± 1.91	23.81 ± 2.52	3.20 ± 0.41	7.44 ± 0.78
S315A	0.79 ± 0.21	11.16 ± 2.83	5.84 ± 0.52	22.78 ± 2.32	5.31 ± 0.52	20.81 ± 2.06	6.78 ± 1.13	1.78 ± 0.30	3.81 ± 0.67
E319A	0.73 ± 0.15	17.04 ± 2.94	10.01 ± 0.72	57.77 ± 4.63	9.14 ± 0.62	52.68 ± 4.23	12.51 ± 1.68	3.08 ± 0.37	4.06 ± 0.52
W412A	0.91 ± 0.22	-	4.63 ± 0.25	-	4.21 ± 0.25	-	4.67 ± 0.67	-	-
A404E	1.86 ± 0.43	22.01 ± 4.84	22.30 ± 1.63	33.71 ± 4.44	13.21 ± 0.94	30.69 ± 3.94	7.11 ± 1.02	4.17 ± 0.68	1.71 ± 0.28
N435A	0.91 ± 0.14	-	7.46 ± 0.42	-	6.84 ± 0.43	-	7.52 ± 0.78	-	-
GLUT5*	-	10.91 ± 1.78	-	52.30 ± 2.90	-	43.28 ± 2.40	-	4.04 ± 0.37	-

The results shown are mean from $n = 3$ biologically independent experiments (values reported are mean ± s.e.m. of the fit). The W412A and N435A mutants abolish D-fructose uptake, and therefore D-fructose kinetics could not be measured (-).

*The D-fructose rat GLUT5 kinetic parameters in proteoliposomes were measured as for PfHT1, using protein purified as previously described⁵.

Extended Data Table 2 | Data collection and refinement statistics (molecular replacement)

Glucose bound	
Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	73.74, 189.45, 136.89
α, β, γ (°)	90.0, 96.4, 90.0
Resolution (Å)	21.85 - 3.65 (3.80-3.65)
R_{merge} (%)	28.5 (>100)
$I/\sigma I$	9.8 (0.51)
CC^*	0.99 (0.57)
Completeness (%)	99.4 (99.7)
Redundancy	12.4 (12.7)
Refinement	
Resolution (Å)	21.85 - 3.65
No. reflections	41238
$R_{\text{work}} / R_{\text{free}}$ (%)	27.3/28.4
No. atoms	
Protein	13908
Ligand/ion	96
Water	-
B-factors	
Protein	230.80
Ligand/ion	196.06
Water	-
R.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.54

Data were obtained by scaling together two datasets that were collected on different crystals. The highest-resolution shell used in the final refinement is shown in parentheses.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Synchrotron software controllers MxCuBE at ESRF and Generic Data Acquisition (GDA) software at Diamond Light Source

Data analysis

Prism 7 - for data plotting and analysis
CCP4 v7 - Crystallographic software suite
Phenix 1.11.1-2575 - Crystallographic software suite
BUSTER 2.10.3 - Crystallographic software suite
PyMol - Molecular graphics software
XDS - Diffraction data processing package
Coot - Crystallographic model building
Gromacs 2018.1 - MD simulation package
MODELLER version 9.21 - MD simulation package

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Structure have been deposited to the PDB with accession code 6RW3 (glucose bound)

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size Individual transport experiments were routinely performed as technical replicates (n=3) to assess standard experimental variations

Data exclusions No data was excluded.

Replication All proteoliposome transport experiments were repeated at least 3 times and the results were reproduced each time.

Randomization Animals or human research participants were not involved in this study and, as such, samples were not randomized for the experiments.

Blinding Animals or human research participants were not involved in this study and, as such, blinding was not used in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a Involved in the study

☐ ☒ Unique biological materials

☒ ☐ Antibodies

☒ ☐ Eukaryotic cell lines

☒ ☐ Palaeontology

☒ ☐ Animals and other organisms

☒ ☐ Human research participants

Methods

n/a Involved in the study

☒ ☐ ChIP-seq

☒ ☐ Flow cytometry

☒ ☐ MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials No restrictions apply for material used in this study, such as plasmids and strains.

Streamflow response to forest management

<https://doi.org/10.1038/s41586-020-1940-6>

Received: 24 July 2019

Accepted: 2 December 2019

Published online: 12 February 2020

James W. Kirchner^{1,2*}, Wouter R. Berghuijs¹, Scott T. Allen^{1,3}, Markus Hrachowitz⁴, Rolf Hut⁴ & Donna M. Rizzo⁵

Arising from: Evaristo, J. & McDonnell, J. J. *Nature* <https://doi.org/10.1038/s41586-019-1306-0> (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).

Forests play a key part in the water cycle, so both planting and removing forests can affect streamflow. In a recent Article¹, Evaristo and McDonnell used a gradient-boosted-tree model to conclude that streamflow response to forest removal is predominantly controlled by the potential water storage in the landscape, and that removing the world's forests would contribute an additional 34,098 km³ yr⁻¹ to streamflow worldwide, nearly doubling global river flow. Here we report several problems with Evaristo and McDonnell's¹ database, their model, and the extrapolation of their results to the continental and global scale. The main results of the paper¹ remain unsubstantiated, because they rely on a database with multiple errors and a model that fails validation tests.

Database problems

We spot-checked the database underlying Evaristo and McDonnell's analysis¹ by comparing individual entries to the original cited references. Roughly half of these spot checks revealed substantial errors in the calculated changes in water yields, or errors in the classification of individual studies as forest planting versus forest removal experiments. Here we describe four examples. (1) The Valtorto catchment in Portugal is classified as a forest clearing experiment¹ although the catchment was never forested, but rather covered by 50-cm-tall heath². The reported post-clearing streamflow increase of 363.6% (ref. ¹) is also inconsistent with table 3 of ref. ², which reports that average streamflow increased by 150%, from 1.0 m³ per day to 2.5 m³ per day. (2) The database reports that forest clearing at the Lemon catchment in Australia increased streamflow by 631.8% (ref. ¹), but from table 1 of ref. ³, we calculate that the average pre- and post-clearing streamflows were 18.0 mm yr⁻¹ and 27.9 mm yr⁻¹ respectively, implying that streamflow increased by only 55%. (3) Brigalow catchments C2 and C3, which each appear twice in the database, are classified as forest planting experiments¹ although neither was planted with forest: C2 was planted with sorghum and wheat and C3 was planted with buffel grass for pasture^{4,5}. (4) Several forest conversion experiments, in which forests were cleared and replanted with other vegetation (for example, references 74, 114, 130 and 163 in ref. ¹), are reported in the database as showing, counterintuitively, large streamflow increases caused by forest planting¹. However, the reported changes in streamflow were calculated relative to intact forest control plots, not cleared land, so they mostly reflect the effects of clearing the existing forest rather than the effects of planting. We suspect that this misattribution of forest clearing effects to forest planting may underlie the paper's surprising finding (see Fig. 2 of ref. ¹ and associated discussion) that forest planting appears to increase streamflow by 100% or more at many sites, with the largest increases at sites with

the highest evapotranspiration rates, a pattern that would normally arise from forest clearing instead.

Model overfitting and validation failure

Gradient-boosted regression trees are data-hungry, and although Evaristo and McDonnell¹ compiled every paired watershed study that they could find, the resulting databases of 161 forest clearing experiments and 90 forest planting experiments are much too small to estimate their seven-variable model reliably. We checked the model codes that Evaristo and McDonnell provided with their paper (see the code availability statement of ref. ¹) and found that the boosted tree algorithm fits 200 free parameters (not counting the dozens of additional free parameters that define the tree's branch points), suggesting substantial overfitting. To test how this overfitting might affect the model's predictions, we split the forest removal and planting databases into training sets (80% of the data) and test sets (the remaining 20% of the data). To balance the distributions of the variables between the training and test sets, we used stratified random sampling; we also used un-stratified random sampling as a more stringent test. We then re-ran the boosted-tree analysis, using the same data, the same platform (JMP, the SAS Institute), and the same algorithm options that Evaristo and McDonnell¹ used, for 300 of these random splits of the data, both with and without 'early stopping' (in which the fitting algorithm stops whenever the next layer would reduce the R^2).

The results in Fig. 1 show that the model fails these validation tests. If the model were not overfitted, the fits to the test data (as measured by the test R^2 on the vertical axis) would be similar to the fits to the training data (as measured by the training R^2 on the horizontal axis), and the dots would lie close to the 1:1 line. Instead, many of the dots lie far below the 1:1 line, and many test R^2 values even lie below zero, indicating model predictions that are worse than random guessing. Figure 1 thus shows that the model is overfitted and makes unreliable predictions (because it is too flexible, and thus has been 'fitted to the noise' in the training data). This result holds whether one uses 'early stopping' or not, and both stratified and un-stratified validation tests yield broadly similar results.

Although individual randomizations can yield test R^2 values that are similar to the training R^2 (or even higher), one should not draw conclusions from such anomalies. Model performance is better reflected in the medians of the training and test R^2 values across many randomization trials (Table 1). Table 1 confirms quantitatively what Fig. 1 shows visually: in each case, the median test R^2 is much smaller than the median training R^2 , and many test R^2 values are below zero.

¹Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland. ²Swiss Federal Research Institute WSL, Birmensdorf, Switzerland. ³Department of Geology and Geophysics, University of Utah, Salt Lake City, UT, USA. ⁴Department of Civil Engineering, Delft University of Technology, Delft, The Netherlands. ⁵Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT, USA. *e-mail: kirchner@ethz.ch

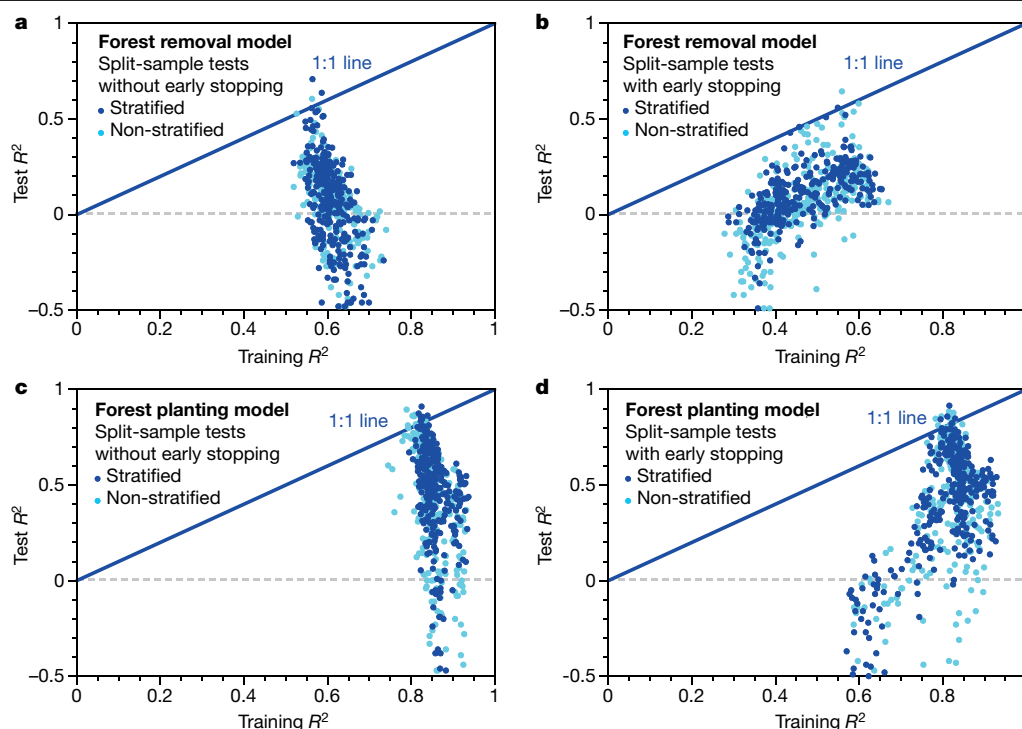


Fig. 1 | Split-sample validation tests of gradient-boosted-tree model fitted to forest clearing and planting data. **a, b,** Model fitted to forest clearing data with and without early stopping; **c, d,** model fitted to forest planting data with and without early stopping. The source data were randomly split into 300 training and test sets in 80/20 ratios, as described in the text. If the model were

not overfitted, the R^2 statistics obtained from the training and test sets would be similar to one another, and thus the dots would lie close to the 1:1 lines. Instead, the test R^2 statistics are generally much smaller than the training R^2 values. Points with test R^2 values less than -0.5 , which indicate that model predictions were much worse than random guessing, are not shown.

All of the paper's¹ main results are based on the boosted-tree model, so the validation failure documented here invalidates the paper's conclusions. The other machine learning methods in the paper have similar validation issues, but we will not explore them in detail because the paper's conclusions do not depend on them.

Exaggerated importance of potential storage

The finding¹ that streamflow response to forest removal was primarily controlled, not by climate, but by total potential water storage in the landscape, was puzzling to us for two reasons. First, it was difficult to imagine how total storage, much of which may lie below the rooting

zone of trees, could be the major control on the hydrological effects of tree removal. Second, given that forest planting and forest removal both alter the same variable (forest cover), but in opposite directions, it was hard to reconcile the paper's two main findings¹: that potential storage is the dominant control on streamflow response to forest clearing (but not planting), and that actual evapotranspiration (AET) is the dominant control on streamflow response to forest planting (but not clearing).

Closer examination reveals that the apparent importance of potential storage relies on one extreme data point (the Lemon catchment, Australia), which has a potential storage of 15 m, more than twice the next-highest value in the dataset. If we remove this one data point, potential storage disappears as the most important factor (Table 2), and is replaced by potential evapotranspiration (PET). This one data point is so influential because Evaristo and McDonnell's analysis¹ uses an 'independent uniform' variable importance profiler. This profiler is intended for use where the likely values of each variable will be uniformly distributed over the range of the data⁶, which is inconsistent with the strongly skewed distributions of potential storage in Evaristo and McDonnell's paired watershed dataset (Fig. 2a) and in their global catchment database (Fig. 2b). Potential storages exceeding 7.5 m comprise only 0.6% of Evaristo and McDonnell's paired watershed dataset (light blue bars, Fig. 2a) and 6% of their global catchment database (light blue bars, Fig. 2b), but 50% of the distribution used to calculate the influence of potential storage, exaggerating its importance.

Although Evaristo and McDonnell fully documented their choice of this "independent uniform" profiler¹, other choices, more consistent with the available data, lead to a different conclusion. For example, if we instead use a profiling method that takes into account the actual distributions of all of the variables ("independent resampled" profiling), PET becomes the most important variable, and potential storage drops to fourth place (Table 2). And if the profiling method also takes account of the correlations among the variables, in addition to their actual

Table 1 | Summary of split-sample validation test results

Model and split-sample test performed (80/20 split in all cases)	Median training R^2	Median test R^2	Fraction of test $R^2 < 0$
Forest removal model			
Stratified, with early stopping	0.449	0.108	31%
Stratified, without early stopping	0.605	0.096	36%
Unstratified, with early stopping	0.458	0.053	34%
Unstratified, without early stopping	0.608	0.057	40%
Forest planting model			
Stratified, with early stopping	0.827	0.455	13%
Stratified, without early stopping	0.852	0.486	10%
Unstratified, with early stopping	0.826	0.475	16%
Unstratified, without early stopping	0.844	0.474	17%

Test results are shown for the boosted-tree model fitted to forest removal and forest planting data. 'Fraction of test $R^2 < 0$ ' indicates the percentage of tests in which model predictions were worse than random guessing.

Matters arising

Table 2 | Relative variable importance using different profilers

Profiling method and treatment of Lemon catchment	Potential evapotranspiration	Runoff coefficient	Drainage area	Potential storage	Actual evapotranspiration	Root zone storage	Permeability
Independent uniform							
Lemon included	0.317 (2)	0.098 (3)	0.036 (5)	0.508 (1)	0.041 (4)	0.007 (6)	0.000 (7)
Lemon omitted	0.500 (1)	0.056 (4)	0.031 (5)	0.299 (2)	0.179 (3)	0.001 (6)	0.001 (6)
Independent resampled							
Lemon included	0.642 (1)	0.114 (3)	0.165 (2)	0.094 (4)	0.030 (5)	0.005 (6)	0.000 (7)
Lemon omitted	0.710 (1)	0.077 (4)	0.134 (2)	0.091 (3)	0.050 (5)	0.001 (6)	0.003 (7)
Dependent resampled							
Lemon included	0.440 (1)	0.189 (2)	0.171 (3)	0.137 (5)	0.109 (6)	0.155 (4)	0.095 (7)
Lemon omitted	0.433 (1)	0.180 (2)	0.174 (3)	0.129 (5)	0.102 (6)	0.161 (4)	0.098 (7)

Relative importance scores for each of the seven variables in Evaristo and McDonnell's forest removal model¹ are shown for three different profiling methods, and for including and excluding the Lemon catchment (see text). Ranks are shown in parentheses. The most important variable in each case is highlighted in bold.

distributions (“dependent resampled” profiling), the most important variable is again PET, and potential storage drops to fifth place out of seven variables (regardless of whether we include or exclude the Lemon catchment; see Table 2).

Exaggerated global streamflow implications

To estimate the potential impact of forest clearing on global streamflow (table 1 of ref. ¹), Evaristo and McDonnell first applied their boosted-tree model to a database of 442,319 catchments for which the required seven input variables are available (whether or not they are actually forested). Evaristo and McDonnell then multiplied the median of the modelled percentage change in streamflow for each continent's catchments by the average continental river flow (see Table 3). Because less than 30% of Earth's land area is forested⁷, however, the potential percentage increase in streamflow from forest clearing should not be applied to the entire continental runoff; that is, one cannot clear forests from the 70% of Earth's land surface where no forests exist. Evaristo and McDonnell's calculation¹ implicitly assumes that Earth's entire land-mass is forested, and leads to unrealistic results. For example, under Evaristo and McDonnell's median scenario¹, their table 1 implies that total post-clearing runoff in Asia would be 95% of total Asian precipitation⁸ (32,140 km³ yr⁻¹; Table 3), a runoff ratio that is rarely observed even in urban areas. For Australia and Oceania, the results in Evaristo and McDonnell's table 1 violate conservation of mass, with total post-clearing runoff (1,970 km³ yr⁻¹ + 5,412 km³ yr⁻¹ = 7,382 km³ yr⁻¹) exceeding total precipitation⁸ (6,405 km³ yr⁻¹).

Distributed over the roughly 40 million square kilometres of the Earth's surface that is actually forested⁷, Evaristo and McDonnell's

claimed global streamflow increase¹ of 34,098 km³ yr⁻¹ implies an average of 850 mm yr⁻¹ more streamflow from cleared forest lands. This value exceeds the streamflow increases that were measured in every one of the 95 paired watershed studies reviewed by Stednick⁹, and exceeds their average by a factor of five.

Back-of-the-envelope calculations suggest different conclusions. Globally, evapotranspiration from forests is roughly 250 mm yr⁻¹ greater than from croplands or grasslands¹⁰, and multiplying this difference by the 40 million square kilometres of global forests⁷ yields a rough estimate of 10,000 km³ yr⁻¹, less than one-third of Evaristo and McDonnell's¹ result. Even this may be an overestimate, because the lower evapotranspiration rates of grasslands partly reflect the fact that they often occur in drier climates; thus the difference between forest and grassland evapotranspiration may exaggerate the effects of converting forests to grasslands.

Concluding remarks

Evaristo and McDonnell are valued colleagues of ours, and we greatly appreciate their transparency in making their data and codes available, without which the issues described here would have been much harder to diagnose. We agree with them that streamflow response to forest management is an important issue that deserves a comprehensive analysis, including subsurface catchment characteristics as potential explanatory variables.

Readers should also keep in mind that this is not a purely academic exercise. How much, and under what conditions, forests should be cleared is an important policy question with wide-ranging consequences for economies, societies and ecosystems. In that regard, we are concerned that the conclusion that “forest removal can lead to increases in streamflow that are around 3.4 times greater than the mean annual

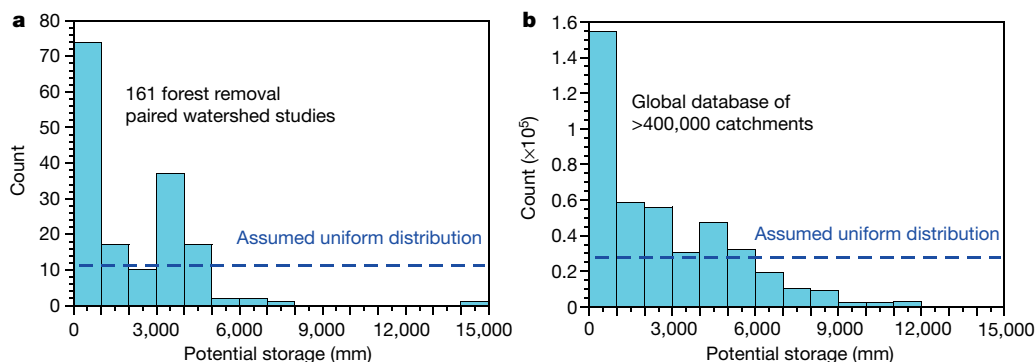


Fig. 2 | Distributions of potential storage, compared to the uniform distribution used to estimate its influence in Evaristo and McDonnell's analysis¹. a, Distribution of potential storage in Evaristo and McDonnell's

dataset of 161 paired watershed studies. **b**, Distribution of potential storage in Evaristo and McDonnell's database of over 400,000 catchments worldwide.

Table 3 | Modelled effects of forest cover change on continental runoff

Region	Total river runoff (km ³ yr ⁻¹) ^a	Change in runoff in response to forest-cover change ^a (km ³ yr ⁻¹)		Total river runoff after removal (km ³ yr ⁻¹) ^b	Total precipitation (km ³ yr ⁻¹) ^c	Change in runoff in response to forest-cover change (%) ^d		Median water yield in complete catchment dataset (%) ^e	
		Planting	Removal			Planting	Removal	Planting	Removal
Africa	4,320	-605(1,944)	8,986(5,616)	13,306	20,780	-14.0(45.0)	208.0(130.0)	-14(45)	208(130)
Asia	14,550	-1,979(5,835)	16,062(25,783)	30,612	32,140	-13.6(40.1)	110.4(177.2)	-14(40)	110(177)
Australia and Oceania	1,970	-412(725)	5,412(4,962)	7,382	6,405	-20.9(36.8)	274.7(251.9)	-21(36)	275(252)
Europe	3,240	-875(1,102)	813(1,426)	4,053	7,165	-27.0(34.0)	25.1(44.0)	-27(34)	25(44)
North and Central America	6,200	-806(2,034)	918(2,102)	7,118	13,910	-13.0(32.8)	14.8(33.9)	-13(33)	15(34)
South America	10,420	0(3,751)	1,908(17,559)	12,328	28,355	0.0(36.0)	18.3(168.5)	0(36)	18(168)
Totals	40,700	-4,676	34,098	74,799	109,755				

Values with parentheses are medians (and interquartile ranges).

^aFrom table 1 of ref. ¹.

^bSum of total river runoff and median change due to removal.

^cTotal precipitation from ref. ⁸, which is also the original source of the total river runoff values.

^dMedian and IQR of runoff changes, as percentage of total river runoff.

^eMedian and IQR of water yield predictions (each rounded to the nearest percentage point in the published database) for Evaristo and McDonnell's 442,319 'complete' catchments. These agree within roundoff error with the percentages calculated by dividing the change in runoff by the total runoff for each continent. This agreement demonstrates that the changes in runoff shown in table 1 of ref. ¹ were calculated by multiplying the median (and IQR) of the percentage water yield predictions by the total river runoff, rather than by the runoff from forested areas.

runoff of the Amazon River¹¹ is overstated and could be misinterpreted. The Amazon flows continuously, but the streamflow benefits of forest clearing are transient, typically lasting only a few years, or at most decades, after felling¹¹. One must also keep in mind that the water transpired by vegetation is an important source of precipitation farther downwind, estimated to account for roughly 40% of continental precipitation¹⁰. Thus, sustained large-scale clearing of forests would predictably lead to precipitation decreases and drying of continental interiors, although the precise magnitude of this effect remains difficult to constrain.

Data availability

All of the data analysed here are available as described in the data availability and code availability statements of ref. ¹, or from the cited references.

1. Evaristo, J. & McDonnell, J. J. Global analysis of streamflow response to forest management. *Nature* **570**, 455–461 (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).
2. Stoof, C. R. *et al.* Hydrological response of a small catchment burned by experimental fire. *Hydrol. Earth Syst. Sci.* **16**, 267–285 (2012).
3. Ruprecht, J. K. & Schofield, N. J. Effects of partial deforestation on hydrology and salinity in high salt storage landscapes. I. Extensive block clearing. *J. Hydrol.* **129**, 19–38 (1991).

4. Lawrence, P. A. & Thorburn, P. J. (eds) *Changes in Hydrology, Soil Fertility, and Productivity of Brigalow Catchments Following Clearing Project Report RQR89002* (Department of Primary Industries, Brisbane, 1989).
5. Thornton, C. M., Cowie, B. A., Freebairn, D. M. & Playford, C. L. The Brigalow catchment study. II. Clearing brigalow (*Acacia harpophylla*) for cropping or pasture increases runoff. *Aust. J. Soil Res.* **45**, 496–511 (2007).
6. SAS Institute Inc. *JMP 14 Profilers* (SAS Institute Inc., 2018).
7. *FAO Global Forest Resource Assessment 2015* (Food and Agriculture Organization of the United Nations, 2016).
8. L'vovich, M. I. & White, G. F. in *The Earth as Transformed by Human Action* (eds Turner, B. L. *et al.*) 235–252 (Cambridge Univ. Press, 1990).
9. Stednick, J. D. Monitoring the effects of timber harvest on annual water yield. *J. Hydrol.* **176**, 79–95 (1996).
10. Ellison, D., Futter, M. N. & Bishop, K. On the forest cover-water yield debate: from demand- to supply-side thinking. *Glob. Change Biol.* **18**, 806–820 (2012).
11. Andréassian, V. Waters and forests: from historical controversy to scientific debate. *J. Hydrol.* **291**, 1–27 (2004).

Author contributions All authors discussed the issues raised here, and contributed to the writing. J.W.K. analysed the data and led the writing effort.

Competing interests The authors declare that they have no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.W.K.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Forest age and water yield

<https://doi.org/10.1038/s41586-020-1941-5>

Received: 28 June 2019

Accepted: 2 December 2019

Published online: 12 February 2020

Adriaan J. Teuling^{1*} & Anne J. Hoek van Dijke^{1,2,3}

Arising from: Evaristo, J. & McDonnell, J. J. *Nature* <https://doi.org/10.1038/s41586-019-1306-0> (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).

Planting and removal of forest affect average streamflow (also referred to as water yield), but there is ongoing debate as to what extent this long-term difference between precipitation and evapotranspiration is modulated by local conditions. A recent paper by Evaristo and McDonnell¹ introduces a conceptual vegetation-to-bedrock model to explain variability in reported streamflow responses to changes in forest cover based on an analysis of seven factors that describe climate, soil properties and catchment size. Their analysis excludes well known controls—such as the percentage of catchment area under change², forest type and time since afforestation—that we show here to be important. By excluding these primary controls, Evaristo and McDonnell risk attributing water yield response to co-varying secondary controls rather than to the underlying causes.

We illustrate the importance of the record length (or time since afforestation) using unique longterm measurements of water yield made under controlled conditions. At Castricum in The Netherlands, and St Arnold in Germany, two large lysimeters were planted with coniferous and deciduous trees in the 1940s and 1960s, respectively,

while reference conditions (bare soil and grassland, respectively) were maintained in an additional lysimeter. At both stations, strong, consistent and continuing declines in average water yield response were observed over averaging periods that ranged from several years up to the whole experiment duration (Fig. 1), coinciding with a steady increase in tree height and biomass^{3,4} and in spite of possible limitations in rooting depth. The declines follow an exponential decay (with a coefficient of determination of 0.91 or larger) with an e-folding time τ of 15 years and a stronger water yield response for coniferous forest than for deciduous forest. As a result, each individual lysimeter already covers a range in water yield response of 30% up to 70%, comparable to the total range reported by Evaristo and McDonnell across different watersheds¹. Similar response times were found for afforestation experiments in deciduous broadleaf forest in North Carolina in the USA⁵ and at the German lysimeter station of Britz-Eberswalde⁶, while analysis of longterm streamflow data in Sweden revealed similar strong effects of forest biomass and age⁷.

The record length of the studies used by Evaristo and McDonnell¹ varies considerably from 1 year to 75 years, but is mostly lower than the timescale of water yield response to forest growth of 15 years (Fig. 1). Therefore, it is likely that the values reported in studies with record lengths of up to once or even twice the e-folding time (15–30 years) are in fact highly sensitive to the length of their record. The mixing of data with variable record lengths could explain why Evaristo and McDonnell

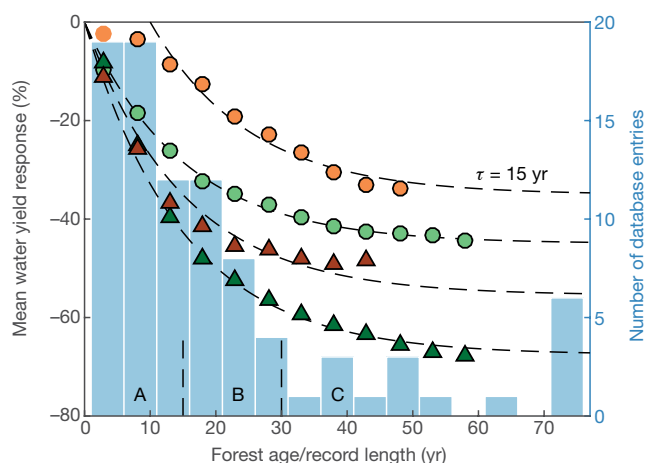


Fig. 1 | Impact of forest age on water yield response to forest planting. Data points are from coniferous (triangles) and deciduous (circles) lysimeters at Castricum (green) and St Arnold (red/orange). Dashed curves indicate exponential fits with a characteristic timescale τ of 15 years, with a 10-year shift assumed for the deciduous lysimeter in St Arnold. Letters A, B and C indicate record length (or forest age) domains used in Fig. 2. The background histogram shows the distribution of the record length of the forest planting studies used by Evaristo and McDonnell. Note that most studies (82%) have a record length of less than 30 years, and strong changes in water yield response are observed in this period. This figure and Fig. 2 were generated by Matlab 2015b (<http://nl.mathworks.com/products/matlab/>).

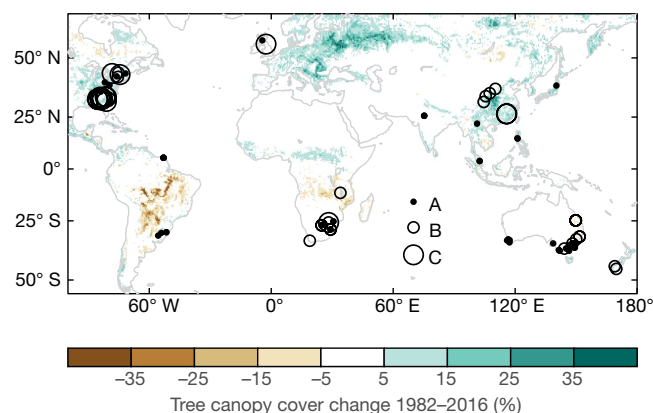


Fig. 2 | Global tree canopy cover change distribution and record length of water yield response to forest planting. Points/circles indicate locations of forest planting studies used by Evaristo and McDonnell¹, with the size reflecting the record length according to classes A, B and C as indicated in Fig. 1. The background map shows changes in tree canopy cover over the period 1982–2016 obtained from a recent analysis of satellite data⁸.

¹Hydrology and Quantitative Water Management Group, Wageningen University and Research, Wageningen, The Netherlands. ²Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, Wageningen, The Netherlands. ³Environmental Sensing and Modelling, Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg. *e-mail: ryan.teuling@wur.nl

find actual evapotranspiration (AET) to be the factor explaining most of the magnitude, rather than timing, of water yield response to planting. When the location of stations with sufficient record length are added to a global map of changes in forest cover over the recent decades⁸, it becomes clear that accurate observations of longterm impacts of forest planting on water yield are concentrated in only a few regions. Strikingly, the forest cover change hotspots are observational blind spots for water cycle impacts. Given the potential of large-scale afforestation to offset carbon emissions⁹, a robust understanding of the hydrological impacts of current and future forest management is more important than ever.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Five-year-average water yield observations used in the analysis are provided in Extended Data Table 1.

1. Evaristo, J. & McDonnell, J. J. Global analysis of streamflow response to forest management. *Nature* **570**, 455–461 (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).

2. Bosch, J. M. & Hewlett, J. D. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *J. Hydrol.* **55**, 3–23 (1982).
3. Harsch, N., Brandenburg, M. & Klemm, O. Large-scale lysimeter site St. Arnold, Germany: analysis of 40 years of precipitation, leachate and evapotranspiration. *Hydrol. Earth Syst. Sci.* **13**, 305–317 (2009).
4. Teuling, A. J. A forest evapotranspiration paradox investigated using lysimeter data. *Vadose Zone J.* **17**, 170031 (2018).
5. Swift, L. W. & Swank, W. T. Long term responses of streamflow following clearcutting and regrowth/Réactions à long terme du débit des cours d'eau après coupe et repeuplement. *Hydrol. Sci. Bull.* **26**, 245–256 (1981).
6. Müller, J. Forestry and water budget of the lowlands in northeast Germany—consequences for the choice of tree species and for forest management. *J. Water Land Dev.* **13a**, 133–148 (2009).
7. Jaramillo, F. et al. Dominant effect of increasing forest biomass on evapotranspiration: interpretations of movement in Budyko space. *Hydrol. Earth Syst. Sci.* **22**, 567–580 (2018).
8. Song, X.-P. et al. Global land change from 1982 to 2016. *Nature* **560**, 639–643 (2018).
9. Griscom, B. W. et al. Natural climate solutions. *Proc. Natl Acad. Sci. USA* **114**, 11645–11650 (2017).

Author contributions A.J.T. conceived the idea and carried out the analysis. A.J.H.v.D. contributed to interpretation of results and writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1941-5>.

Correspondence and requests for materials should be addressed to A.J.T.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Matters arising

Extended Data Table 1 | Observed water yield at long-term lysimeter stations

Site	Period	P	Reference	Broadleaf	Deciduous
St. Arnold	1966-1970	932.04	496.4	484.02	441.2
St. Arnold	1971-1975	677.28	357.24	340.44	191.8
St. Arnold	1976-1980	676.94	346.38	271.86	127.54
St. Arnold	1981-1985	773.34	439.62	334.94	198.8
St. Arnold	1986-1990	791.16	442.96	252.58	173.78
St. Arnold	1991-1995	872.9	530.08	328.16	276.72
St. Arnold	1996-2000	813.24	376.56	181.38	140.42
St. Arnold	2001-2005	835.98	391.74	153.74	171.48
St. Arnold	2006-2010	799.86	333.92	133.68	141.1
St. Arnold	2011-2013	703.43	253.57	130.6	NaN
Castricum	1941-1945	790.2	590.4	533.2	540.8
Castricum	1946-1950	791.4	596.4	433.4	351
Castricum	1951-1955	835.4	631.4	374	208.2
Castricum	1956-1960	857.6	664.4	339.4	190.6
Castricum	1961-1965	873.4	663.6	367.4	204
Castricum	1966-1970	910.8	700.2	366	175.6
Castricum	1971-1975	762.2	546	230.6	87.75
Castricum	1976-1980	783.6	597	270	122.8
Castricum	1981-1985	891.8	682	341.2	164.4
Castricum	1986-1990	848.8	657.2	361.2	118.4
Castricum	1991-1995	933.8	735.6	378.6	147.2
Castricum	1996-1997	744	550	145.5	2

Precipitation data are shown as reference. The reference lysimeter is grassland at St Arnold and bare soil at Castricum. Data after 2007 were not considered for the lysimeter with deciduous forest at St Arnold owing to storm damage caused by cyclone Kyrill. All units are millimetres per year.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No data was collected for this study

Data analysis Graphs were produced in MATLAB

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Five-year average water yield observations used in the analysis are provided in Extended Data Table 1.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Comment on previous paper, with simple analysis of decades-old data that has been used in many previous studies
Research sample	Data comes from fixed lysimeters stations, one for each land cover type
Sampling strategy	No sampling involved
Data collection	AT the stations, observations have been done continuously for decades
Timing and spatial scale	Stations have operated for decades, size of each lysimeter approximately 400 m2
Data exclusions	Data after 2007 were not considered for the lysimeter with deciduous forest at St. Arnold due to storm damage caused by Cyclone Kyrill.
Reproducibility	n/a
Randomization	n/a
Blinding	n/a
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

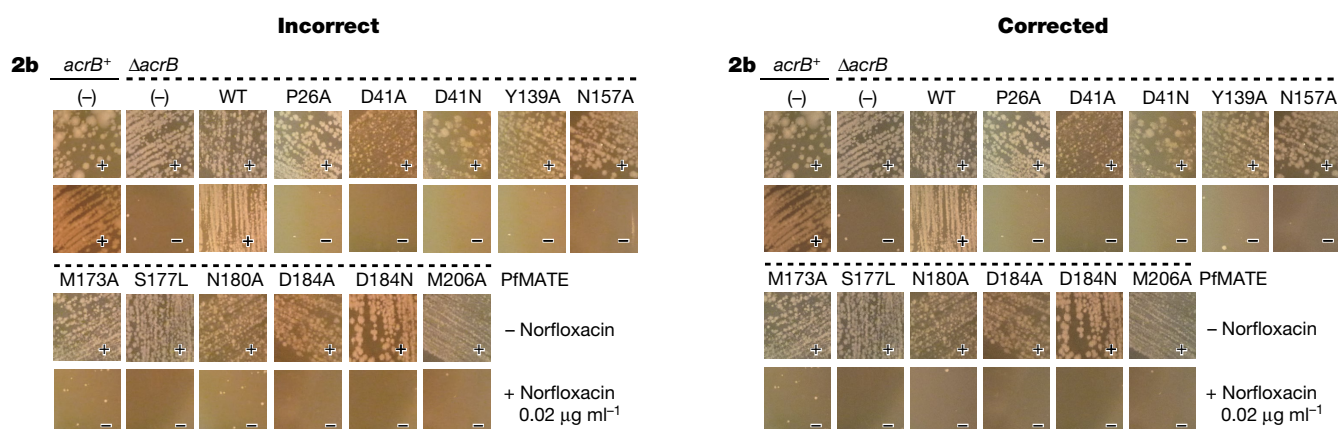
n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Published online 27 March 2013

In this Article, three of the images in Fig. 2b are incorrect. In the growth complementation tests of $\Delta acrB$ strains, all 14 lower panels, from (–) for +Norflaxin to M206A for +Norfloxacin, should have been cropped from the original plate images. However, during the preparation of Fig. 2b we inadvertently cropped the lower panels for Y139A, N157A and N180A from the wrong plates (P26A, M206A and M173A, respectively). Figure 1 shows the incorrect, as-published original Fig. 2b and the corrected Fig. 2b, with the three affected lower panels (Y139A, N157A and N180A for +Norflaxin) now corrected. Since all the growth scores for both the corrected and the incorrect, as-published panels are negative (that is, marked ‘–’ for not complemented), these errors do not change our conclusion for the evaluation of the mutation effects. The original Article has not been corrected online.



Nature | Vol 578 | 13 February 2020 | E19

Publisher Correction: Recycling lithium-ion batteries from electric vehicles

<https://doi.org/10.1038/s41586-019-1862-3>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1682-5>

Published online 06 November 2019

Gavin Harper, Roberto Sommerville, Emma Kendrick, Laura Driscoll, Peter Slater, Rustam Stolkin, Allan Walton, Paul Christensen, Oliver Heidrich, Simon Lambert, Andrew Abbott, Karl Ryder, Linda Gaines & Paul Anderson

In this Review Article, owing to a mistake in renumbering, there were several errors in the author affiliations in the HTML version. The PDF and print versions were correct. The errors have been corrected online.

Retraction Note: Global analysis of streamflow response to forest management

<https://doi.org/10.1038/s41586-020-1945-1>

Retraction to: *Nature* <https://doi.org/10.1038/s41586-019-1306-0>, published online 17 June 2019; corrected online 30 September 2019; addendum 30 September 2019

Jaivime Evaristo & Jeffrey J. McDonnell

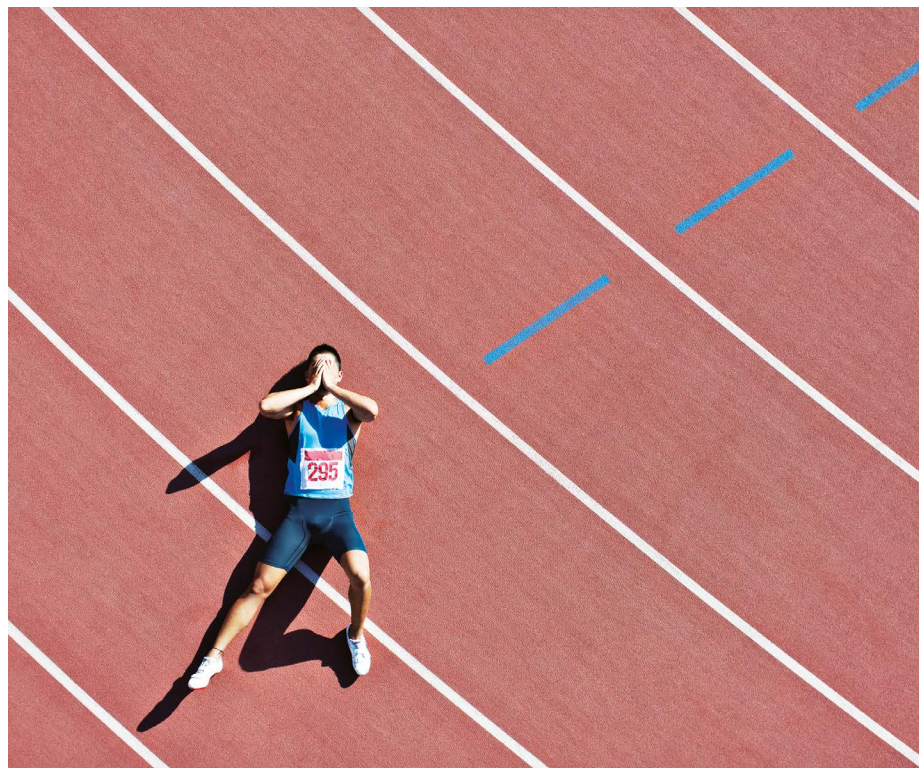
A few weeks after publication of this Article, as a result of comments from James Kirchner and colleagues, we realized that our assembled dataset of paired watershed studies, used to assess the streamflow response to forest removal and planting, contains errors in the percentage change in streamflow associated with land cover modifications. Second, the effects of continent-wide forest removal on streamflow (Table 1 of the Article) are overestimated, because we assumed a starting condition of 100% forest cover. Third, there are serious concerns regarding model validation that need to be assessed using the corrected data. Correcting these honest mistakes goes beyond a simple Author Correction, and therefore we and the *Nature* editors wish to retract this Article. There are two Matters Arising that accompany this Retraction Note, by James W. Kirchner et al. (<https://doi.org/10.1038/s41586-020-1940-6>) and by Adriaan J. Teuling & Anne J. Hoek van Dijke (<https://doi.org/10.1038/s41586-020-1941-5>). We have decided not to respond because we do not wish to cause further confusion by defending a retracted paper. We are working with Kirchner and colleagues to constructively address the issues raised in a revised paper; if and when it is published we will alert readers by posting a comment to this Retraction.

Correspondence should be addressed to J.J.M.

Work

Your
story

Send your careers story
to: naturecareerseditor@nature.com



Sports and science careers might be vastly different — but both can trigger an identity crisis.

AN ACADEMIC IDENTITY CRISIS

Overdoing PhD work can lead to loss of identity. Three things help recover it. **By Robert Seaborne**

During the final 18 months of my PhD programme, I became incredibly absorbed in my work. For months on end, I could be found toiling in the laboratory or writing in an office for 13–14 hours per day. Evenings and weekends that I once spent playing football, going to the gym or socializing were instead used to work on my experiments, read, write or analyse data. I became obsessed with my project. Every waking moment was spent furthering my studies. Every conversation I had revolved around my work. I had become the living embodiment of my PhD, and completely lost my sense of self. I had assumed a new identity: one that centred on my degree programme.

Identity crises are neither a new nor a

unique phenomenon. Elite athletes, for example, are particularly susceptible to them¹, and these events have severe psychological and performance-related effects. It's easy to imagine why: the life of an athlete is the relentless pursuit of perfection in an extremely volatile environment. That promotes extreme dedication, and a win-at-all-costs mentality.

Research suggests that athletes who

“Over time, I have slowly started to gain back an identity that I once lost to my PhD.”

identify entirely as athletes, as opposed to those who see being an athlete as only a facet of their personality, are at greater risk of mental-health damage when this identity is challenged, under threat¹ or removed entirely. These individuals have effectively built an entire identity around one component of their being. And when this identity is challenged or becomes strained, the individual perceives the threat as an attack or criticism of their entire person, leaving them psychologically and emotionally fragile. This is most strikingly seen in elite athletes who are forced to retire; this process effectively strips them of the one identity they have associated with for many years².

Elite sport and academia might seem like two completely distant worlds, but I think they are similar when it comes to their ability to trigger an identity crisis. Both are highly intensive, performance-driven, turbulent careers, with too many candidates trying to ‘make it’ compared with the number of places available.

My own identity had become entirely defined by my PhD work, and I had created a personality defined by just one aspect of my life. When this was under threat and challenged by poor results or failed experiments, I interpreted these outcomes as evidence that my entire identity was a failure or was insufficient. Consequently, my emotional and psychological outlook ebbed and flowed to the rhythm of my PhD. During the highs, I was motivated, excited and passionate about life. But during the lows, I became irritable, aggressive and both physically and mentally drained. I was unstable and unhappy.

I graduated towards the end of 2018, and it has taken me a full year to truly discover, understand and reflect on what this identity crisis was, how it affected me and what mechanisms helped me to overcome it. Identifying and developing these coping strategies was crucial, and would have served me very well had I been advised of these tactics early in my studies. Here I describe three mechanisms that worked for me, in the hope that they might benefit those who are currently in, or who might encounter, a similar scenario.

Exercise

Sport has always been a huge part of my life, but was something that I had lost during the intense periods of my PhD programme. Following the successful defence of my dissertation, I suddenly had a lot of spare time at weekends and evenings. So I decided to restart my outdoor exercise habits. I joined

a local football team and a gym, and I began recreationally going rock climbing and playing tennis. Committing to exercise and competitive sport again has helped me to have another element of my life to focus on outside academia. It gives me a lot of perspective, and helps me to counterbalance the challenges I face during my research career.

Sleep

During the most intense periods of my PhD programme, I prioritized my work over everything else – including getting enough sleep. Your mind works in a much more efficient and productive manner if you are getting sufficient amounts of quality sleep. With this comes a better ability to interpret, process and deal with challenges at both the emotional and psychological level.

Reading

As researchers, we tend to be inquisitive and eager to learn. I realized that if I was to try to resolve my psychological state, then I needed to understand the issue. And so, I read. I read books about how to control the mind^{3,4} through to ones about the habits of highly successful chief executives⁵, businesses⁶ and past and present sporting greats^{7,8}. They helped me to learn a little about how the mind works, and how I can better control my own.

As a result, I slowly began to feel more at ease with my thought processes, and began to understand more about who I was. Over time, I have slowly started to gain back an identity that I once lost to my PhD.

Maintaining your personal identity in a career that is highly volatile, stressful and intense is difficult, and your sense of self can so easily be lost. However, it is crucial to differentiate yourself from your work in order to maintain both your mental and physical health. It is important to understand that successes and failures in your research career do not and should not define who you are. You are a person long before you're a PhD researcher.

Robert Seaborne is a postdoctoral researcher at Queen Mary University of London.
e-mail: r.seaborne@qmul.ac.uk

1. Brewer, B. A., Van Raalte, J. L. & Linder, D. E. *Int. J. Sport Psychol.* **24**, 237–254 (1993).
2. Wyllie, P., Alfermann, D. & Lavallee, D. *Psychol. Sport Exercise* **5**, 7–20 (2004).
3. Dweck, C. *Mindset: The New Psychology of Success* (Ballantine Books, 2007).
4. Peters, S. *The Chimp Paradox: The Acclaimed Mind Management Programme to Help You Achieve Success, Confidence and Happiness* (Ebury Digital, 2012).
5. Buffet, W. *The Snowball: Warren Buffett and the Business of Life* (Bantam Books, 2008).
6. Schmidt, E. & Rosenberg, J. *How Google Works* (Grand Central Publishing, 2017).
7. Walsh, B. *The Score Takes Care of Itself: My Philosophy of Leadership* (Portfolio, 2009).
8. Syed, M. *Bounce: The Myth of Talent and the Power of Practice* (Fourth Estate, 2010).

MEN SELF-HYPE THEIR PAPERS

Sensationalistic words attract citations – and men more often use them. **By Chris Woolston**

A language analysis of titles and abstracts in more than 100,000 scientific articles found that papers with both first and last authors who were women were about 12% less likely than male-authored papers to include sensationalistic terms such as 'unprecedented', 'novel', 'excellent' or 'remarkable'. The study, published in *The BMJ*¹, also found that papers missing such words garnered significantly fewer citations.

Researchers tracked 25 positive terms in clinical-research articles published between 2002 and 2017, and input the authors' names into the Genderize database to predict their genders. The team then created models that compared the citation rates and word choice of articles published in the same journals in the same year with the same subject keywords.

The articles in each comparison were presumably of similar quality, but those that had positive words in their title or abstract

"Is language a mirror of society, or does it shape society?"

garnered 9% more citations overall, and 13% more citations in high-impact journals.

The relative reluctance of female authors to use self-flattering words could contribute to a gender gap in citations and impact, says lead author Marc Lerchenmueller, an economist at the University of Mannheim in Germany and the Yale School of Management in New Haven, Connecticut. In the big picture, he adds, these results should encourage scientific authors and editors to think about word choice and its effects. "Scientists should discuss whether using such sales terms is a disservice to the scientific enterprise," he says.

An increasing practice

The discussion seems to be becoming more important: the analysis also found that such self-flattering words were 80% more common in 2017 than they were in 2002. Lerchenmueller notes that this time period marked an explosion in the number of published articles. "Authors are trying to present

research as favourably as possible to attract attention," he says.

At this point, it's impossible to pinpoint exactly why male and female authors would take a different approach to promotional language, Lerchenmueller adds. He points to decades of studies suggesting women are more likely than men to face a backlash from peers and society when they stray beyond stereotypical norms. Women who have been chastised in the past for being too forceful or boastful might edit themselves and tone down their language, he says. Sensationalistic words could also be added or removed at some point during the editorial process – and Lerchenmueller thinks that this possibility warrants closer examination.

The impact of words

This relative lack of inflated language in female-authored papers echoes a 2019 experimental study published by the National Bureau of Economic Research², showing that women gave themselves relatively poor marks in interviews, performance reviews, job applications and other settings. "We found a large and robust gender gap in self-promotion," says Christine Exley, who is a business-administration researcher at Harvard Business School in Boston, Massachusetts. In one measure, women were less likely to describe their performance favourably when selecting from a list of potential adjectives that ranged from 'terrible' to 'excellent'. Exley notes that in an experimental setting, women should have felt no fear of backlash for over-hyping themselves – but the gender gap still persisted.

Lerchenmueller feels that his study touches on some important philosophical questions about the power and meaning of words. "Is language a mirror of society, or does it shape society?" In the world of science, he says, language seems to both reflect and promote bias – and female researchers are facing the consequences.

Chris Woolston is a freelance writer in Billings, Montana.

1. Lerchenmueller, M. J., Sorenson, O. & Jena, A. B. *Br. Med. J.* **367**, 16573 (2019).
2. Exley, C. L. & Kessler, J. B. *NBER working paper 26345* (2019).



Where I work Anne-Marie Imafidon

Photographed for *Nature* by
Leonora Saunders.

Eat. Sleep. STEM. Repeat: these words on my T-shirt are the mantra of Stemettes, a UK-based outreach enterprise I co-founded in 2013 that encourages girls and young women to enter careers in science, technology, engineering and mathematics (STEM).

I'm a computer scientist and have worked for firms including Goldman Sachs, Deutsche Bank and Hewlett-Packard. But I decided to launch this business because I wanted to have a wider impact. If I can inspire more girls to build something for themselves, that's even more important than me making another algorithm or widget.

I especially love events like the one pictured here, when I and 7 other 'Stemettes' spent the day with 200 girls between the ages of 15 and 19. We gathered at G-Research, a data and technology company in central London that hires people with PhDs to work with maths and algorithms. The girls spoke to the audience about things they're passionate about, and had mock interviews with company employees. This all builds confidence,

and you can tell when girls have a eureka moment. That's what we're there for.

We also run events with banks, energy companies and the UK National Health Service. The girls see that all sorts of people work at such places, including women like me. I don't have to be super corporate or change my hair or the way I speak to do my job. I'm wearing trainers. I can be authentic, and they can, too. We show the girls that you don't have to be a maths genius to work in tech. Digital literacy shouldn't be elitist.

Nearly everyone who signs up for our events, whether through school or individually, is female or non-binary. It's something we mandate – girls tend to be more open without a bunch of teenage boys around. Technology hasn't always had the positive impact on the world, the workforce or our daily lives that it could have. Maybe it's because we don't always have the right people in the room.

Anne-Marie Imafidon is the co-founder and chief executive of Stemettes in London, UK.
Interview by Chris Woolston.